

# Using a Cellular Automata Urban Growth Model to Estimate the Completeness of an Aggregated Road Dataset

Tiernan Erickson  
**GEO Division and  
Penn State University**



# Agenda

- Background and Project Motivation
- Urban Modeling with SLEUTH
- Study Method
- Case Study Examples
  - Texas
  - South Dakota
  - Arizona
- Lessons Learned
- Future Directions



# Background

## Address Canvassing:

- Census workers compare what they see on the ground to what is shown on the Census Bureau's address list.
- Next, the census workers will verify, update, or delete addresses already on the list, and add addresses that are missing from the list.
- At the same time, they will also update maps so they accurately reflect what is on the ground.
- Housing unit addresses verified: 145 million
- Census workers hired for address canvassing: 140,000

Source: U.S. Census Bureau. *Address Canvassing Facts/Statistics*. Retrieved June 16, 2012, from <http://2010.census.gov/news/press-kits/one-year-out/address-canvassing/address-canvassing-facts-statistics.html>



# Background

## Geographic Support System (GSS) Initiative:

- Integrated program in support of the 2020 Census:
- Improved address coverage
- Continual spatial feature updates
- Enhanced quality assessment and measurement
- A targeted address canvassing operation during 2019 in preparation for the 2020 Census.
- Collaboration with federal, state, local, and tribal governments and other stakeholders to establish an acceptable address list for each geographic entity.

Source: U.S. Census Bureau. *Geographic Support System (GSS) Initiative*.

Retrieved June 16, 2012, from <http://www.census.gov/geo/www/gss/index.html>





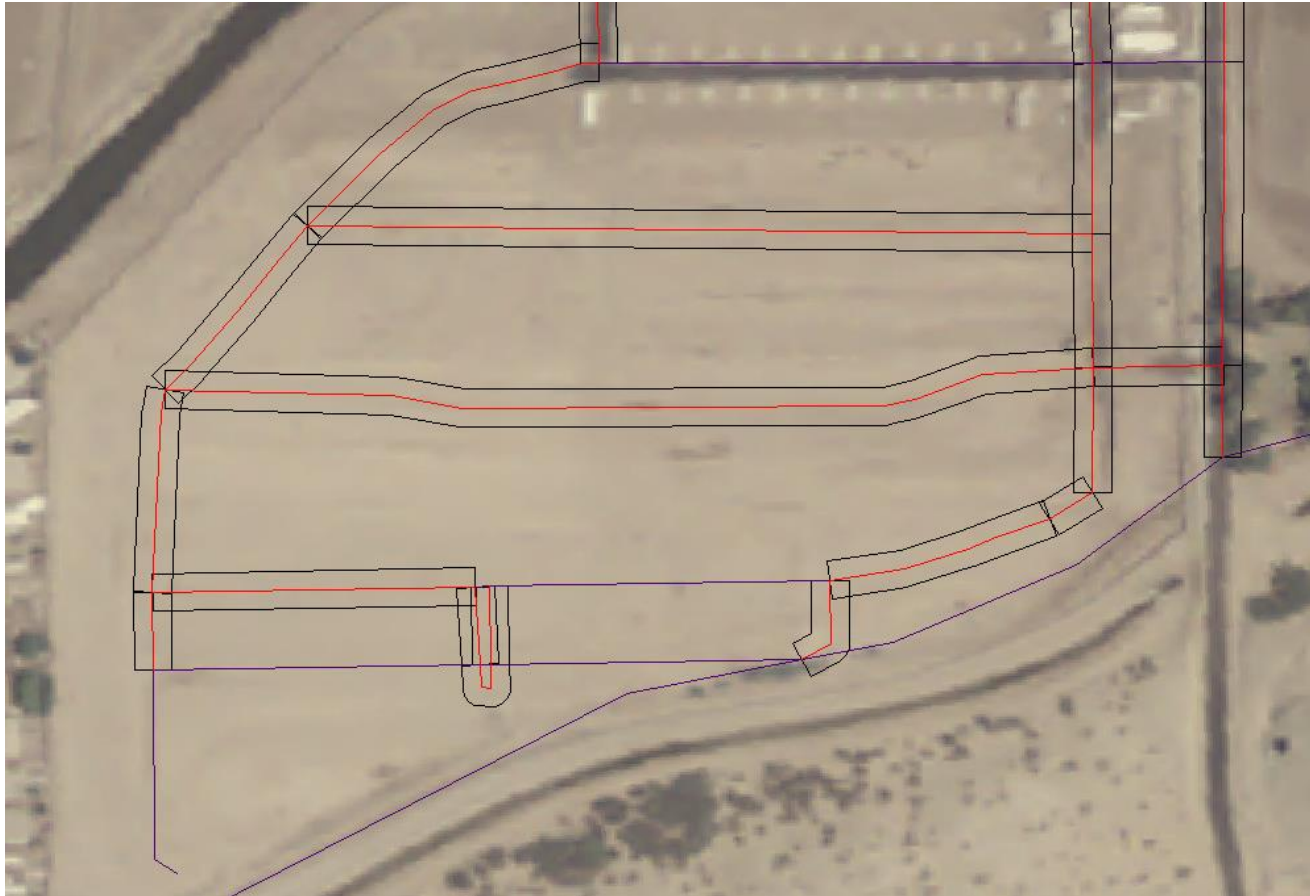
# Background

## Geographic Support System (GSS) Initiative:

- ✓ Positional Accuracy
- ✓ Thematic Accuracy
- ✓ Temporal Accuracy
- ✓ Logical Consistency
- Completeness?



# Spatial Data *Completeness*



# Spatial Data *Completeness*



Detroit, MI

Source: Google Maps





# Spatial Data Completeness



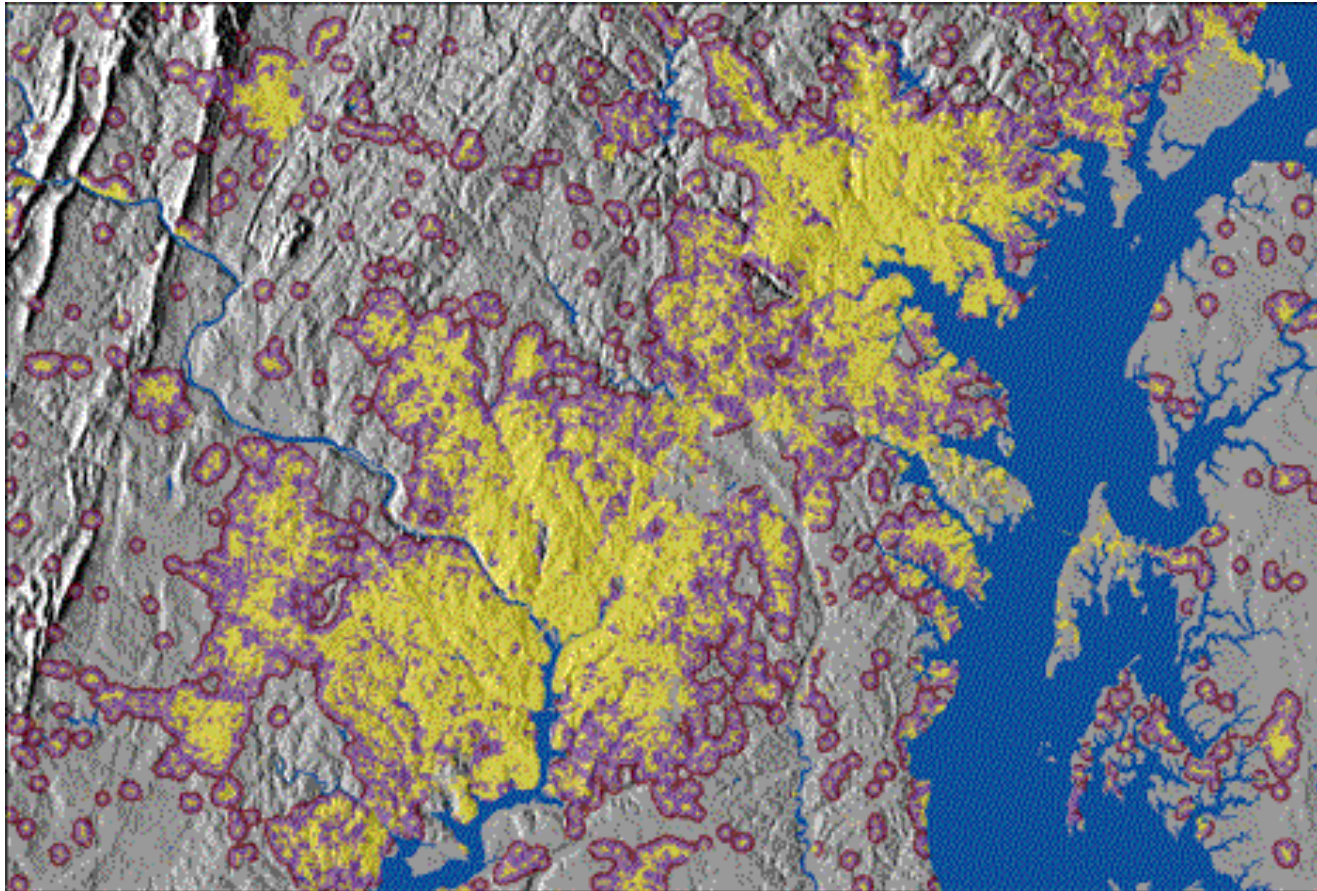
South of Austin, TX

Source: Google Maps





# Urban Growth Forecasting Models



Source: Project Gigalopolis

[http://www.ncgia.ucsb.edu/projects/gig/v2/About/ablImages/apps/wash-balt\\_1792-2100.htm](http://www.ncgia.ucsb.edu/projects/gig/v2/About/ablImages/apps/wash-balt_1792-2100.htm)



# Urban Growth Forecasting Models

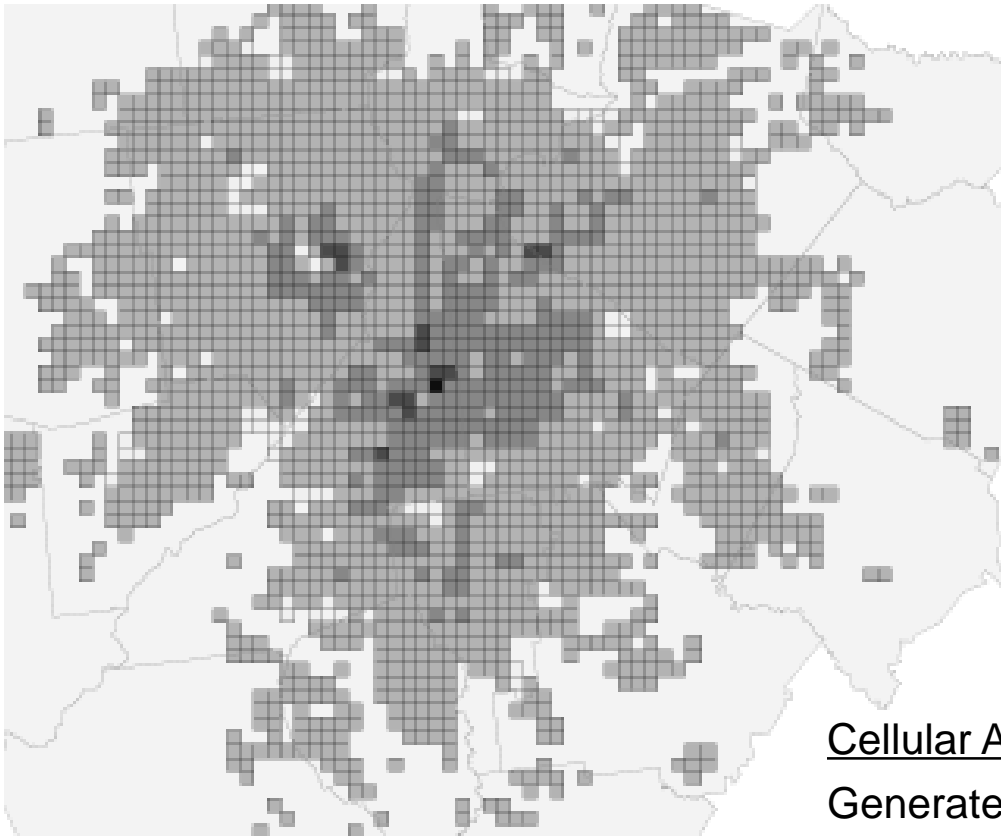


Image Source: Cutsinger and Galster (2006)

## Cellular Automata Urban Growth Models

Generate realistic urban patterns

Integrate the modeling of the spatial and temporal dimensions of urban processes.

-Santé, et al. (2010)





# Urban Growth Forecasting Models

Characteristics for Comparing Urban Growth Models (adapted from Santé, et al., 2010)			
Characteristic	Model Types within Characteristic	Ideal Types for this Project	SLEUTH Fits Ideal Type?
Objective	Descriptive, Predictive, Prescriptive	Predictive	✓ - SLEUTH is designed to predict growth year-by-year.
Cell Space	10m – 500m, also Cadastral Parcels	10m ~ 100m. Finer resolution results would be better for predicting new roads, but have a high computing cost.	✓ - Cell size is variable. Slope and Land Cover data available at 30m resolution. 30-meter cells used in numerous applications.
States	Urban/Non-Urban, Multiple Land Uses Types	Any that allow for exclusion of undevelopable lands: water, preserves, military, etc.	✓ - Uses an Exclusion layer. Parameters allow partial exclusion or attraction of specific areas.
Constraint	Output can be constrained to fit: Annual Growth Rate, Population Growth Projection, Urban Planning Regulations, others	Population Growth Projection would be readily available and make use of accepted Census methods. The Census produces projections by state.	✗ - SLEUTH extrapolates from previous growth in realistic ways, but not constrained to match other models' predictions.
Calibration	Numerous methods in two categories: Trial and Error, Statistical Techniques	The complexity of many models requires trial and error methods.	✓ - SLEUTH uses trial and error methods. Number of trials needed for calibration studied in literature.
Validation	Numerous Methods are proposed for comparing simulated/actual: Ratio of Simulated/Actual Cell Count, Percentage of Correctly Classified Pixels, Urban/Non-Urban Edges Count, Confusion Matrix and Kappa Index, and many others.	Percentage of Correctly Classified Pixels Excluding Original Urban Pixels' intuitively seems like the most sensible metric from among those mentioned in the review. Probably many are suitable.	✓ - SLEUTH uses several measures of goodness-of-fit, based on logistic regression. SLEUTH-3r uses different measures. No consensus on the best to use. Not clear whether users can choose.



# SLEUTH Model

**S - Slope**

**L - Landuse**

**E - Exclusion**

**U - Urban Extent**

**T - Transportation**

**H - Hillshade**



# SLEUTH Model

S - Slope

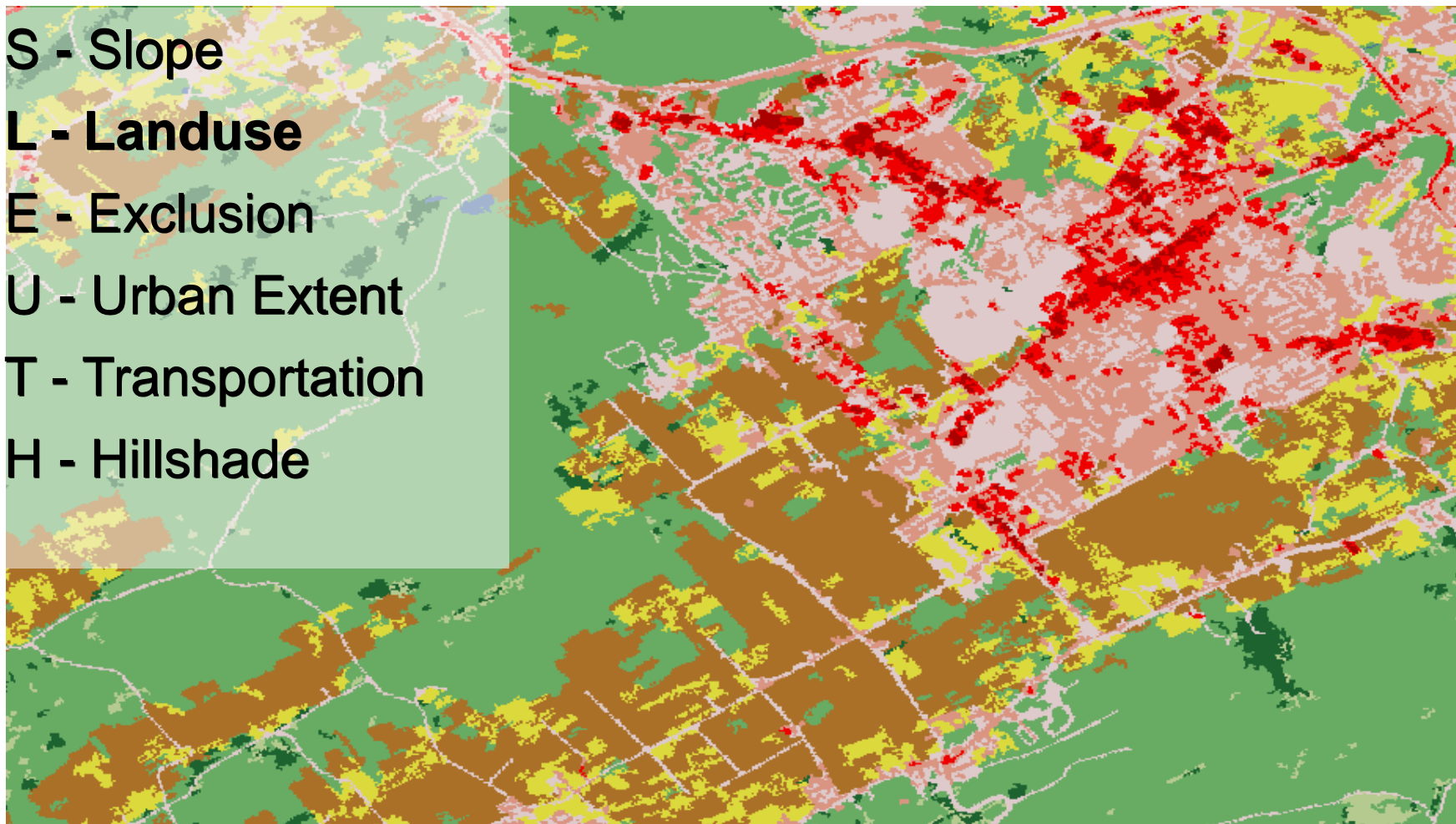
L - Landuse

E - Exclusion

U - Urban Extent

T - Transportation

H - Hillshade





# SLEUTH Model

S - Slope

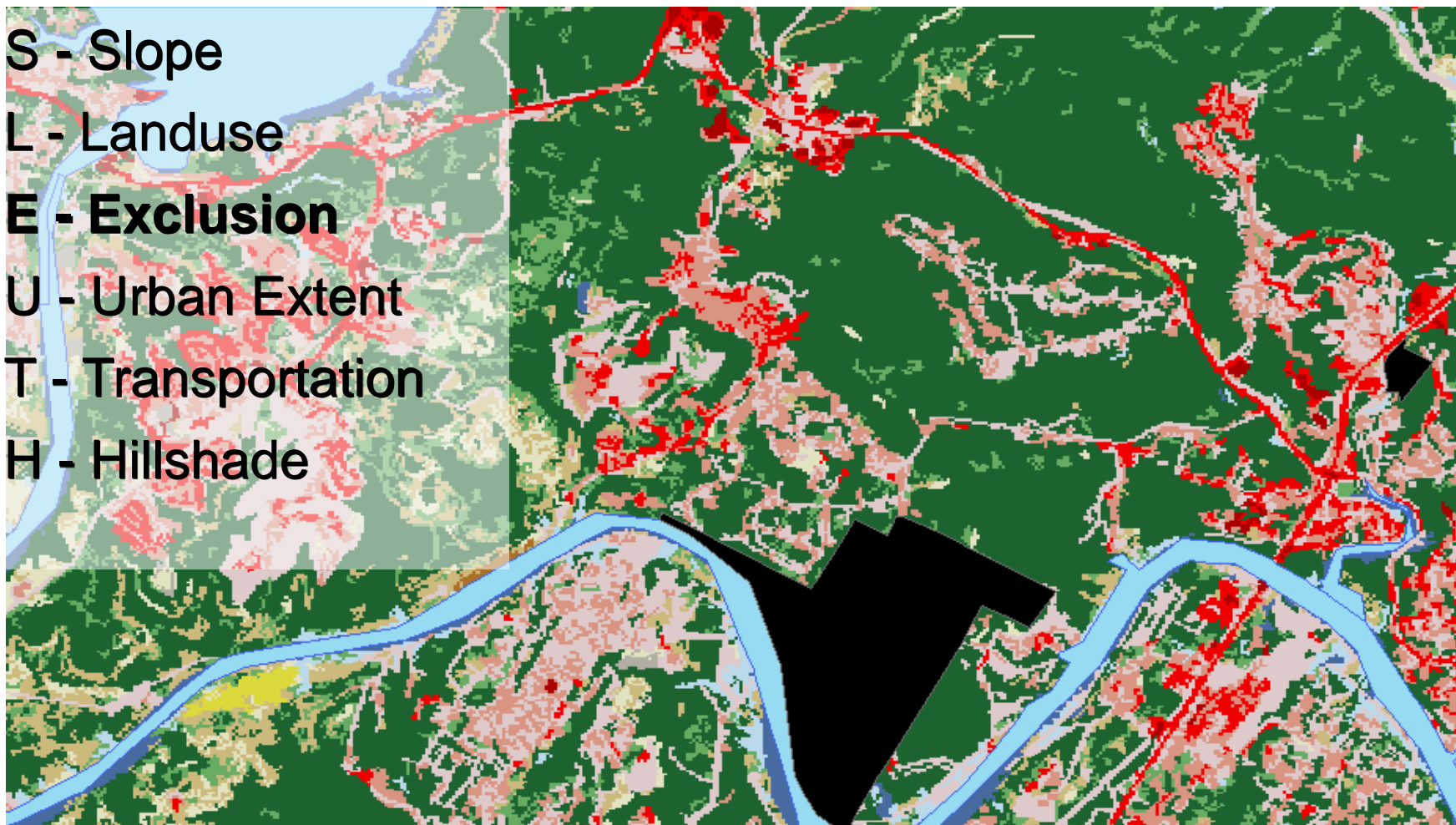
L - Landuse

**E - Exclusion**

U - Urban Extent

T - Transportation

H - Hillshade



# SLEUTH Model

S - Slope

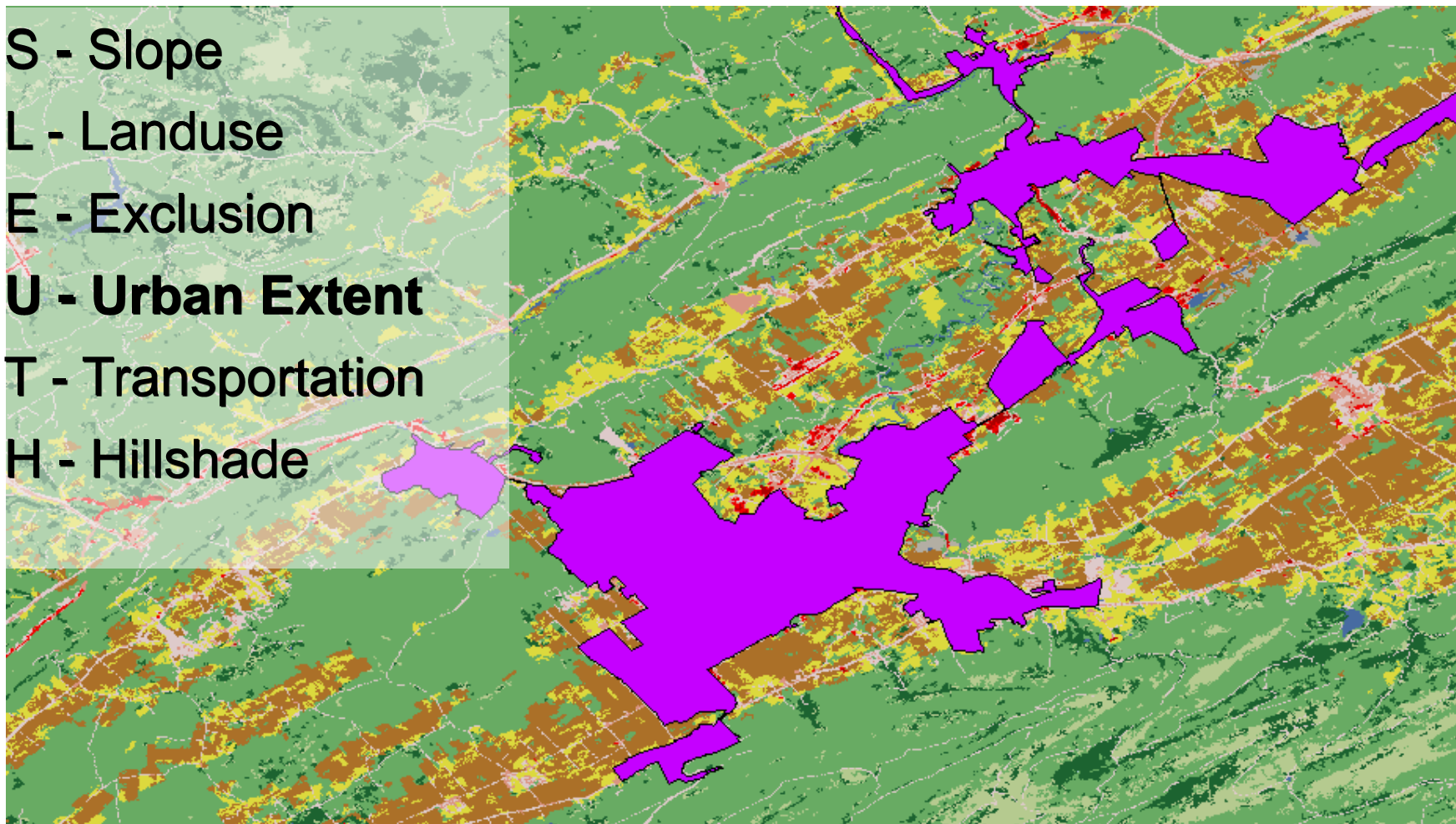
L - Landuse

E - Exclusion

**U - Urban Extent**

T - Transportation

H - Hillshade





# SLEUTH Model

S - Slope

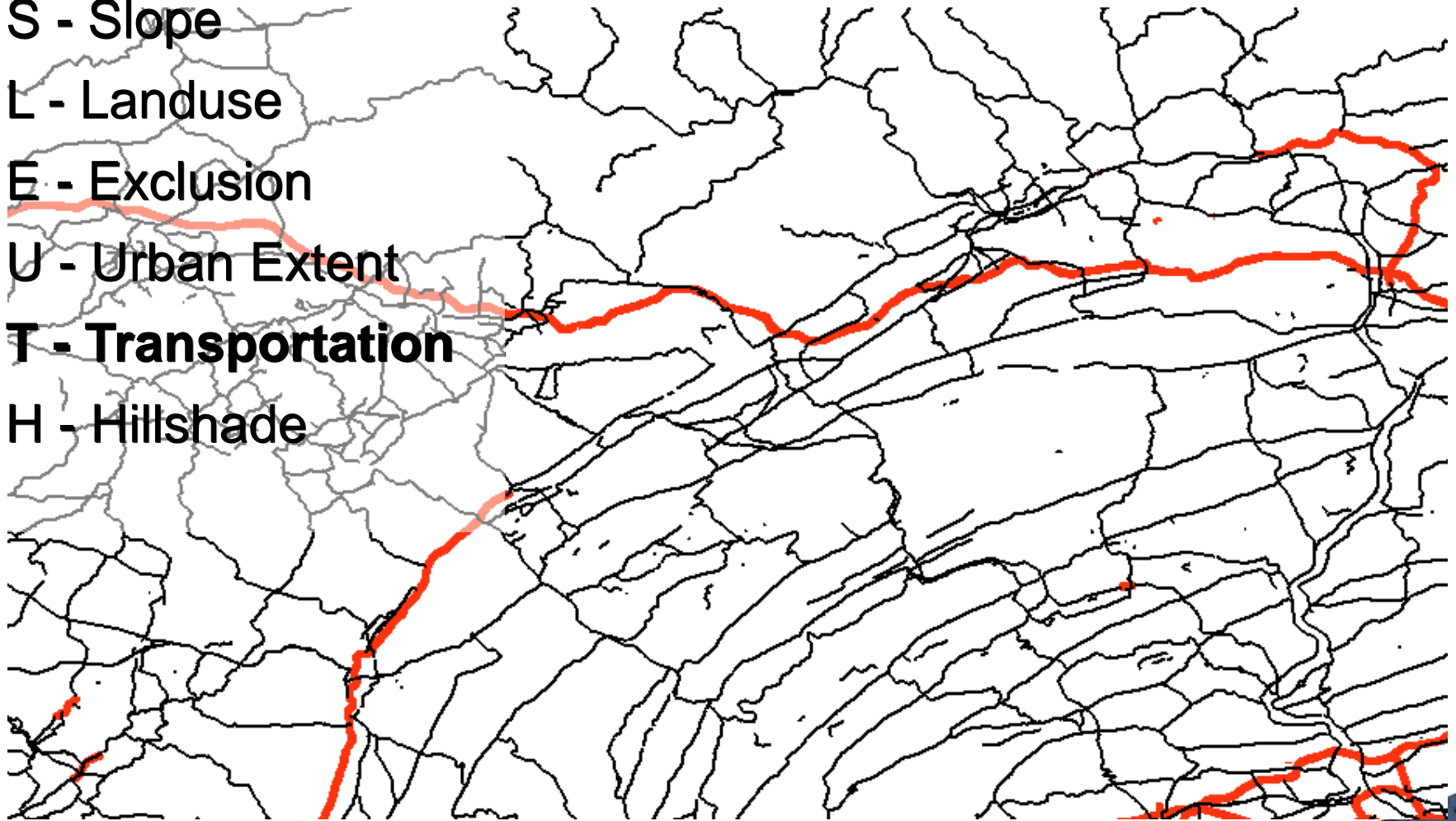
L - Landuse

E - Exclusion

U - Urban Extent

**T - Transportation**

H - Hillshade





# SLEUTH Model

S - Slope

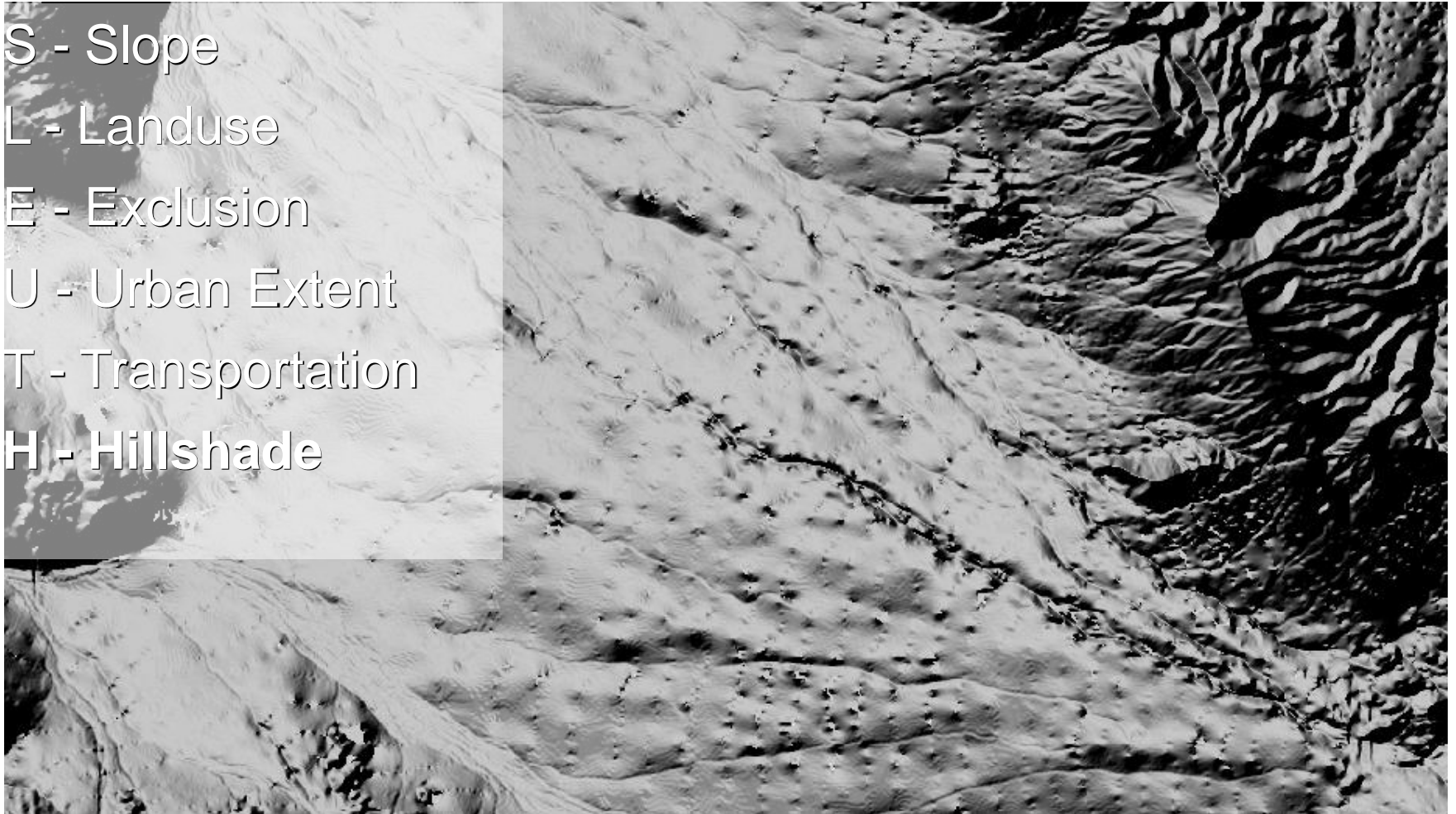
L - Landuse

E - Exclusion

U - Urban Extent

T - Transportation

H - Hillshade



# SLEUTH Model

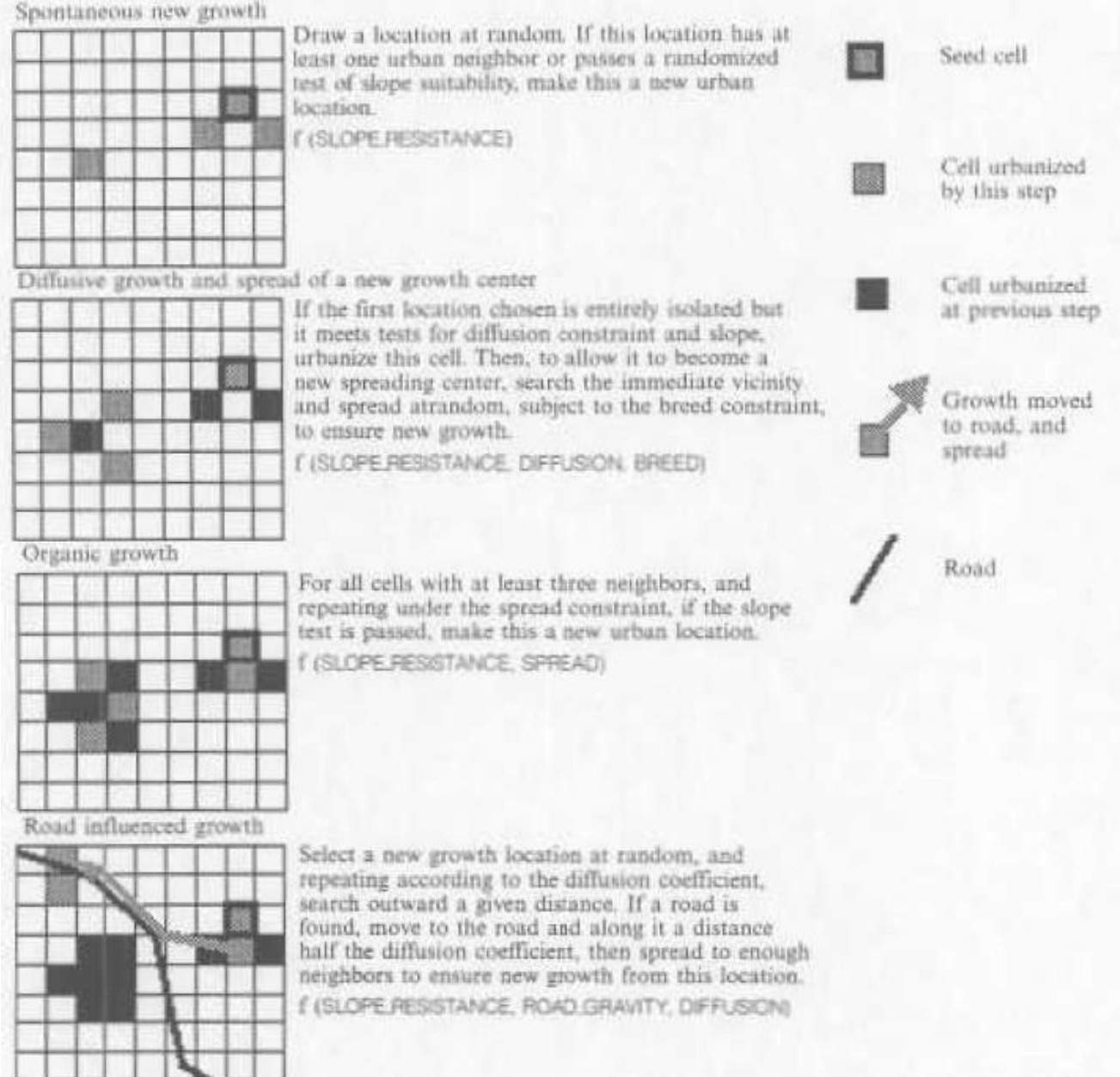
## Model Parameters ("Urban DNA"):

- Diffusion
- Breed
- Spread
- Slope Resistance
- Road Gravity



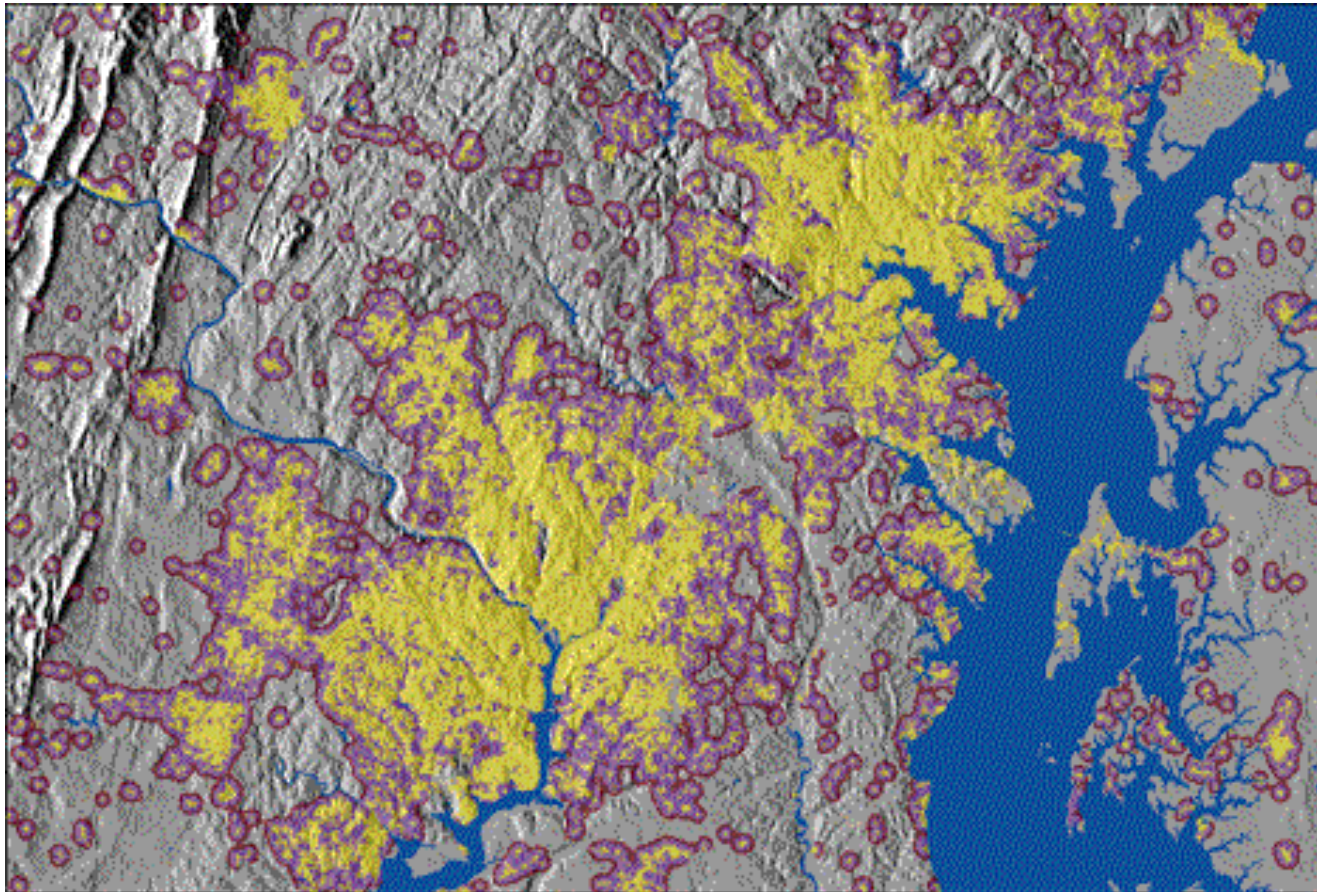
# SLEUTH Growth Types

Source: Clarke et al. (1997)





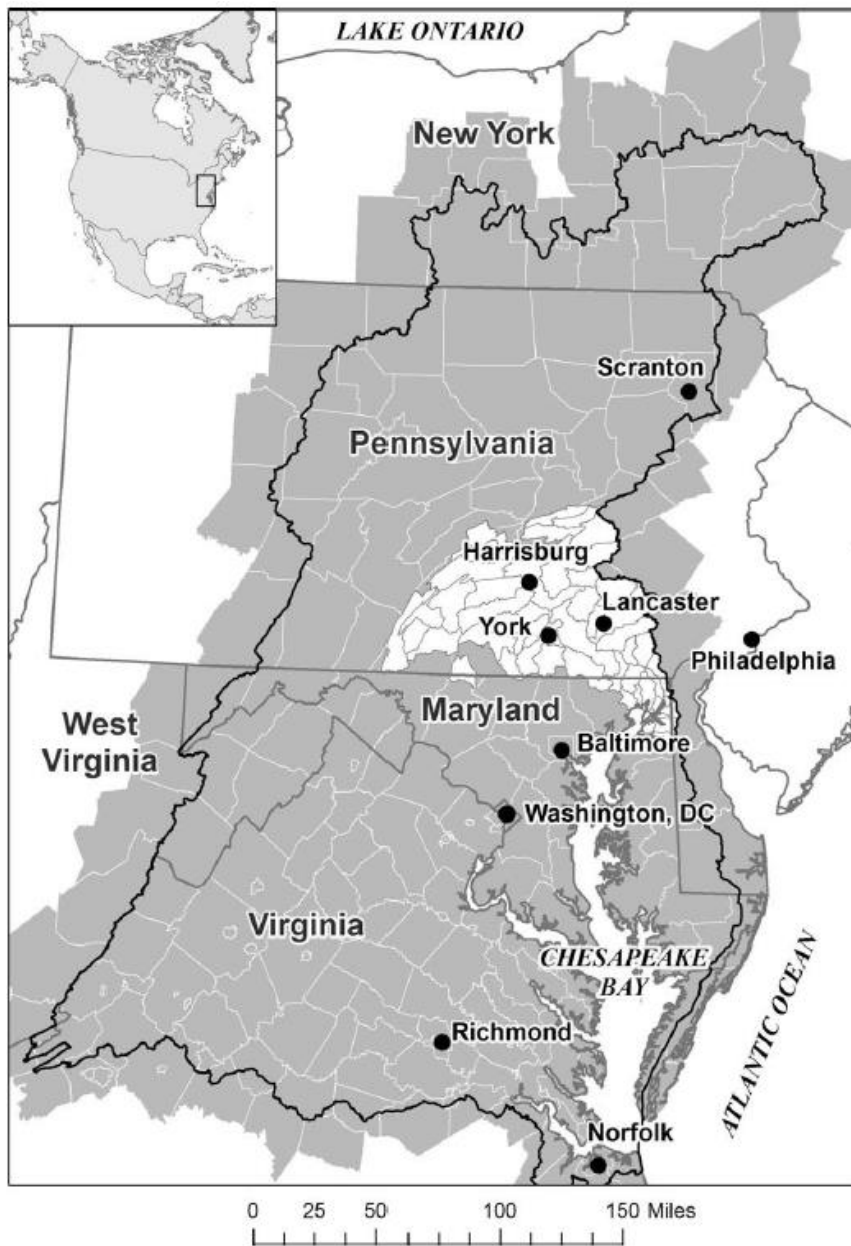
# SLEUTH Model



Source: Project Gigalopolis

[http://www.ncgia.ucsb.edu/projects/gig/v2/About/ablimages/apps/wash-balt\\_1792-2100.htm](http://www.ncgia.ucsb.edu/projects/gig/v2/About/ablimages/apps/wash-balt_1792-2100.htm)





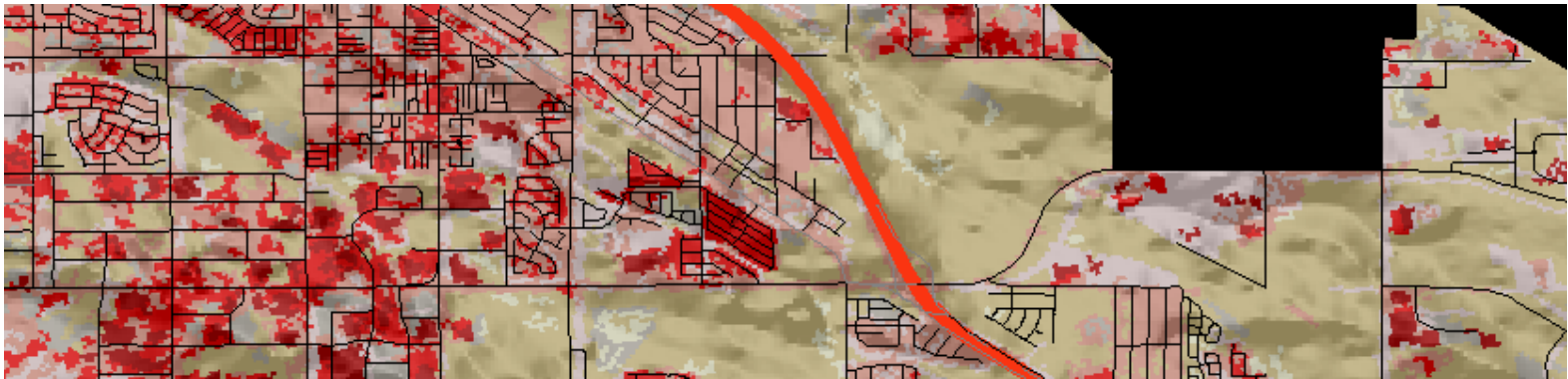
# Method

- Used SLEUTH-3r
- Linux on PC (Cygwin)
- NLCD available for 1992, 2001, and 2006.
- Calibration:
  - 2001 – 2006
- Prediction:
  - 2011(est.)
- Validation:
  - 2011(est.) vs. 2011 (actual)





# Method



- Used SLEUTH-3r
- Linux on PC (Cygwin)
- NLCD available for 1992, 2001, and 2006.
- Calibration: 2001 – 2006
- Prediction: 2011(est.)
- Validation: 2011(est.) vs. 2011 (actual)



# Model Inputs



Model Input Datasets (S.L.E.U.T.H)		Notes
Slope	Digital Elevation Model (DEM), 1 arc-second, 30m cells	National Elevation Dataset (need separate tiles for each area)
Land Cover	National Land Cover Dataset (NLCD), 1 arc-sec., 30m cells	2001, 2006 (nationwide coverage)
Exclusion	TIGER/Line (Area Hydro, Military, National Parks)	2000*, 2006 (need separate files for each county) *2001 TIGER not available
Urbanization	National Land Cover Dataset	
Transportation Network	TIGER/Line (Roads)	
Hillshade	Derived from DEM (for visualization purposes)	National Elevation Dataset (need separate tiles for each area)



# Model Output

## **SLEUTH's Output:**

- Rasters showing percent likelihood of new development for each cell, between 2006 and 2011

## **Research Product:**

- Aggregate prediction values to the tract level.
- Compare predictions at tract level to actual miles of new roads in each tract, 2006 – 2011.
- Generate an estimate for new road growth by 2013, by tract.



# Significance

**Estimate of completeness of aggregated road dataset (TIGER)**

**Incomplete in areas where:**

- 1) Road growth is occurring rapidly, and
- 2) Have not been updated recently

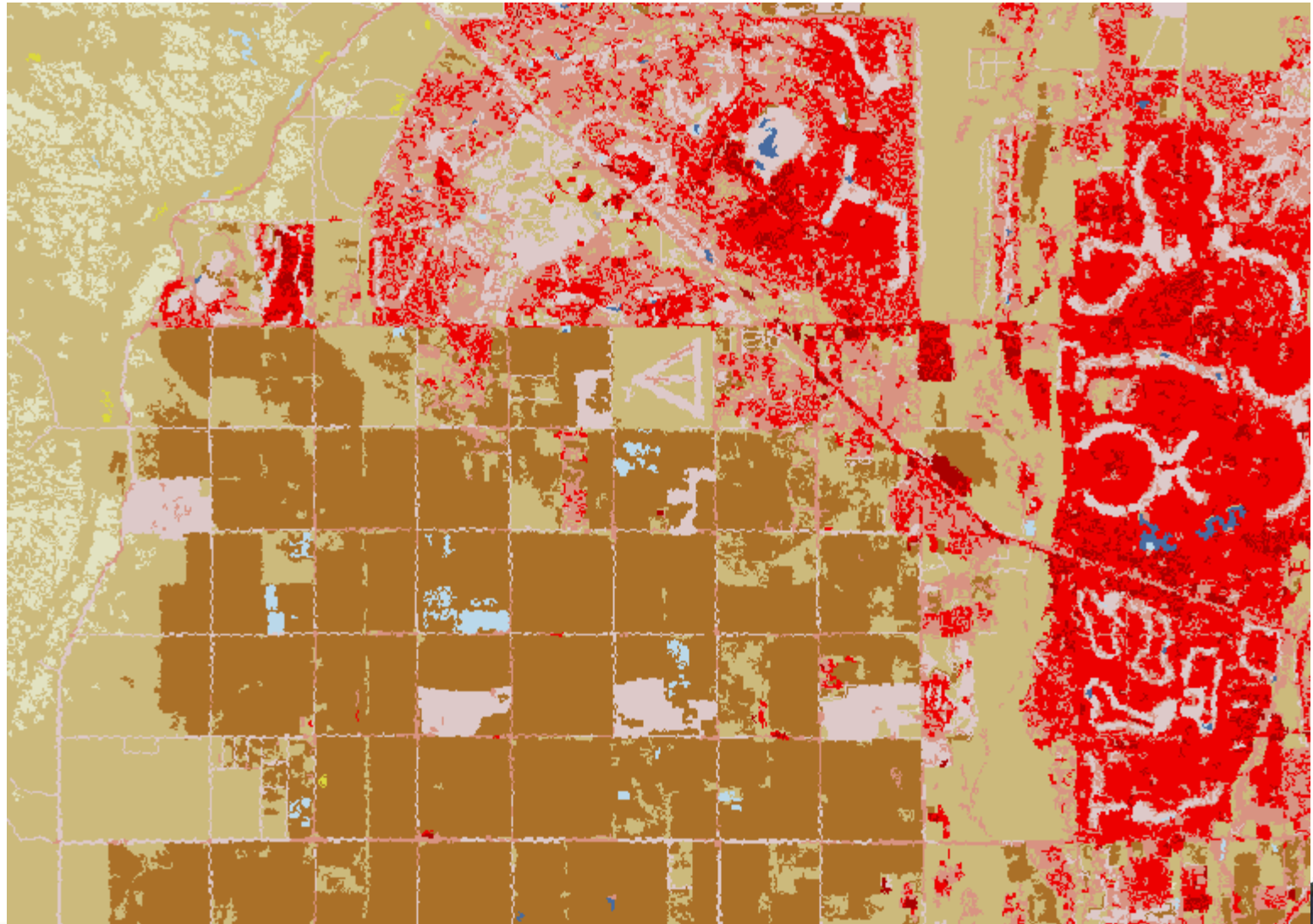
**Complete (save resources) in areas where:**

- 1) Little or no growth, or
- 2) Have been updated recently





# 2001 NLCD



1:100,000; 1 pixel = 30m



# 2001 Urban Landcover 30m cells



1:100,000; 1 pixel = 30m



# 2001 TIGER Roads 2001 Urban Landcover



1:100,000; 1 pixel = 30m





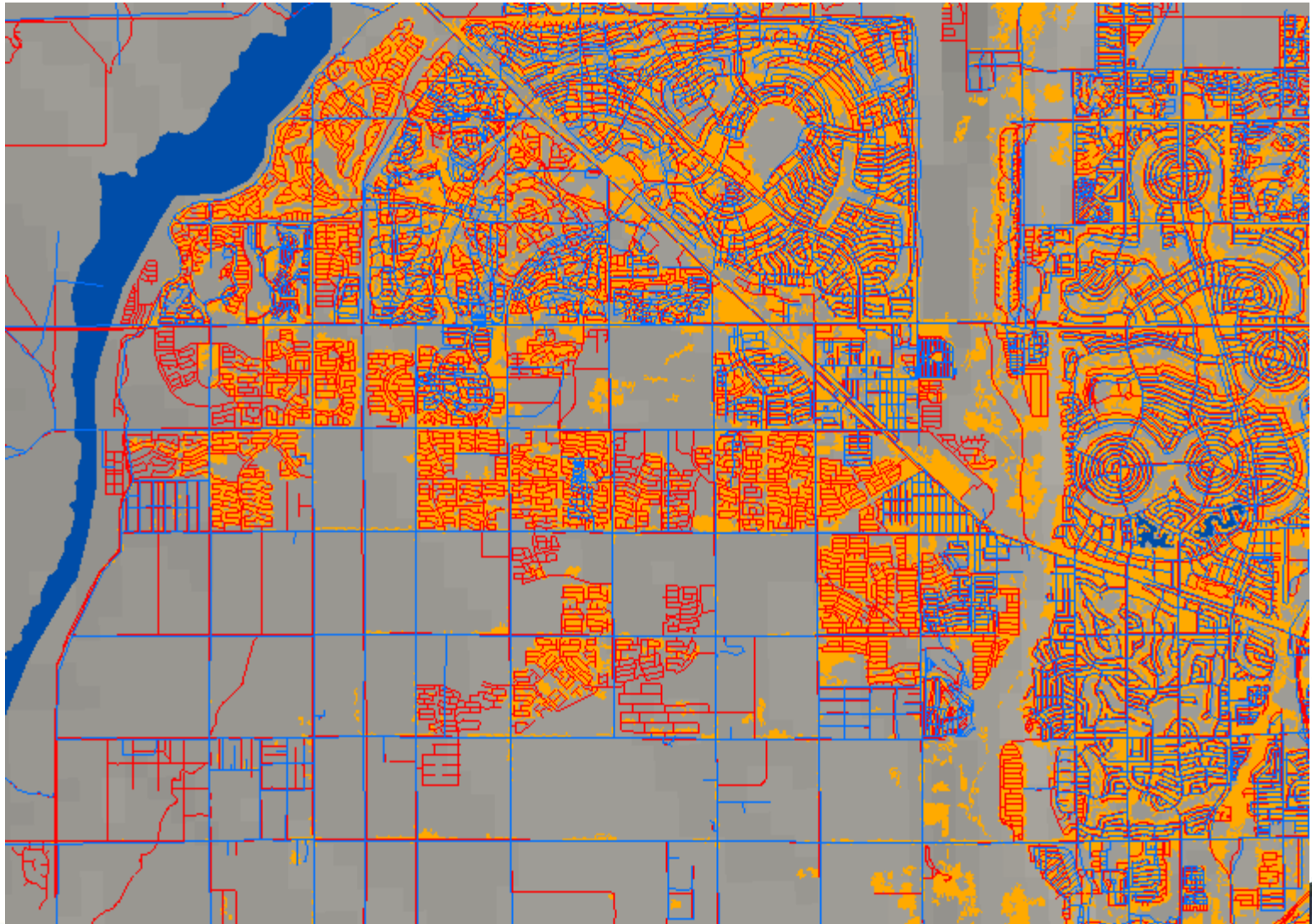
# 2001 TIGER Roads 2006 Urban Landcover



1:100,000; 1 pixel = 30m



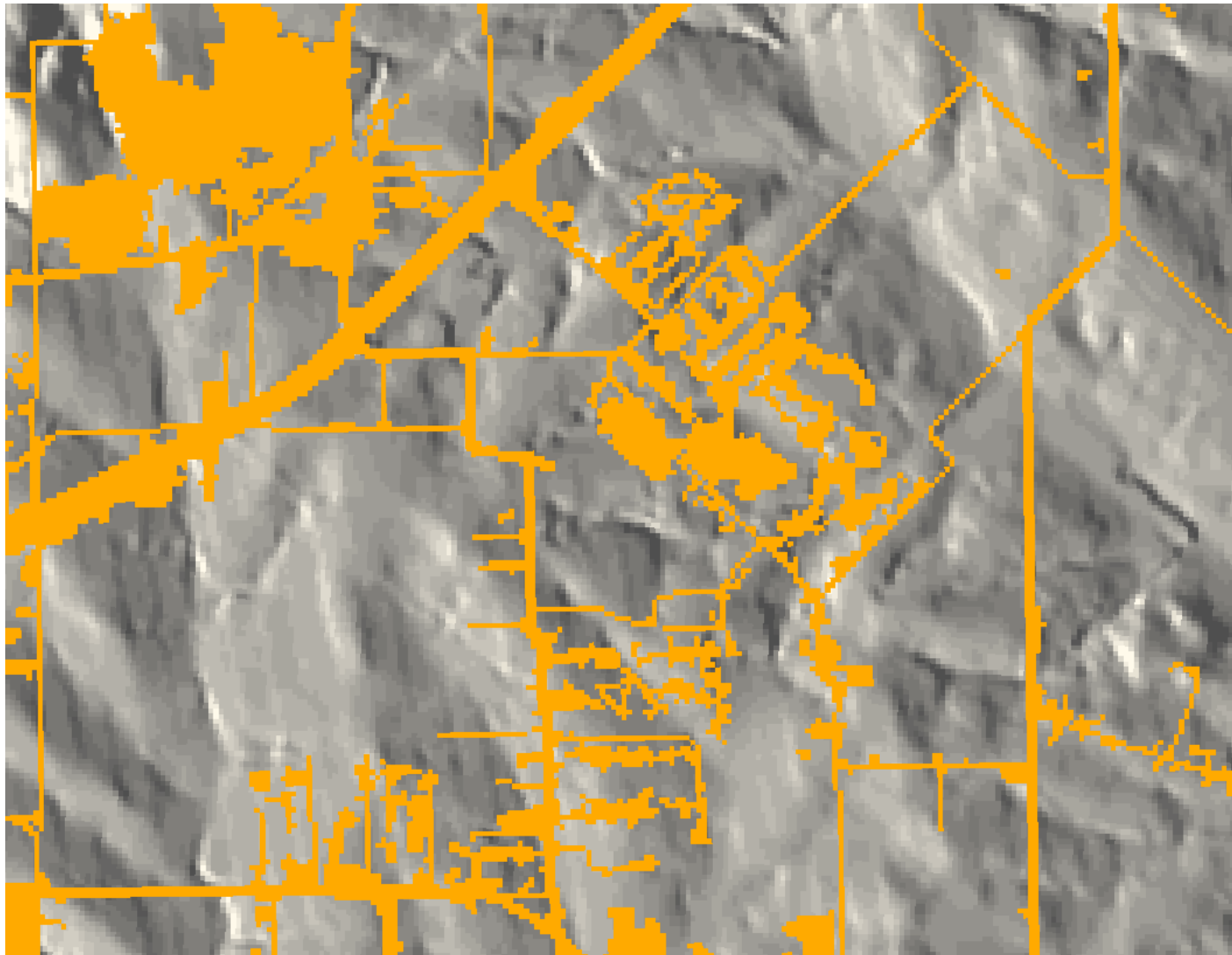
# 2001 Roads (BLUE) 2006 Roads (RED)



1:100,000; 1 pixel = 30m



# Urban Landcover 30m cells



1:20,000; 1 pixel = 30m

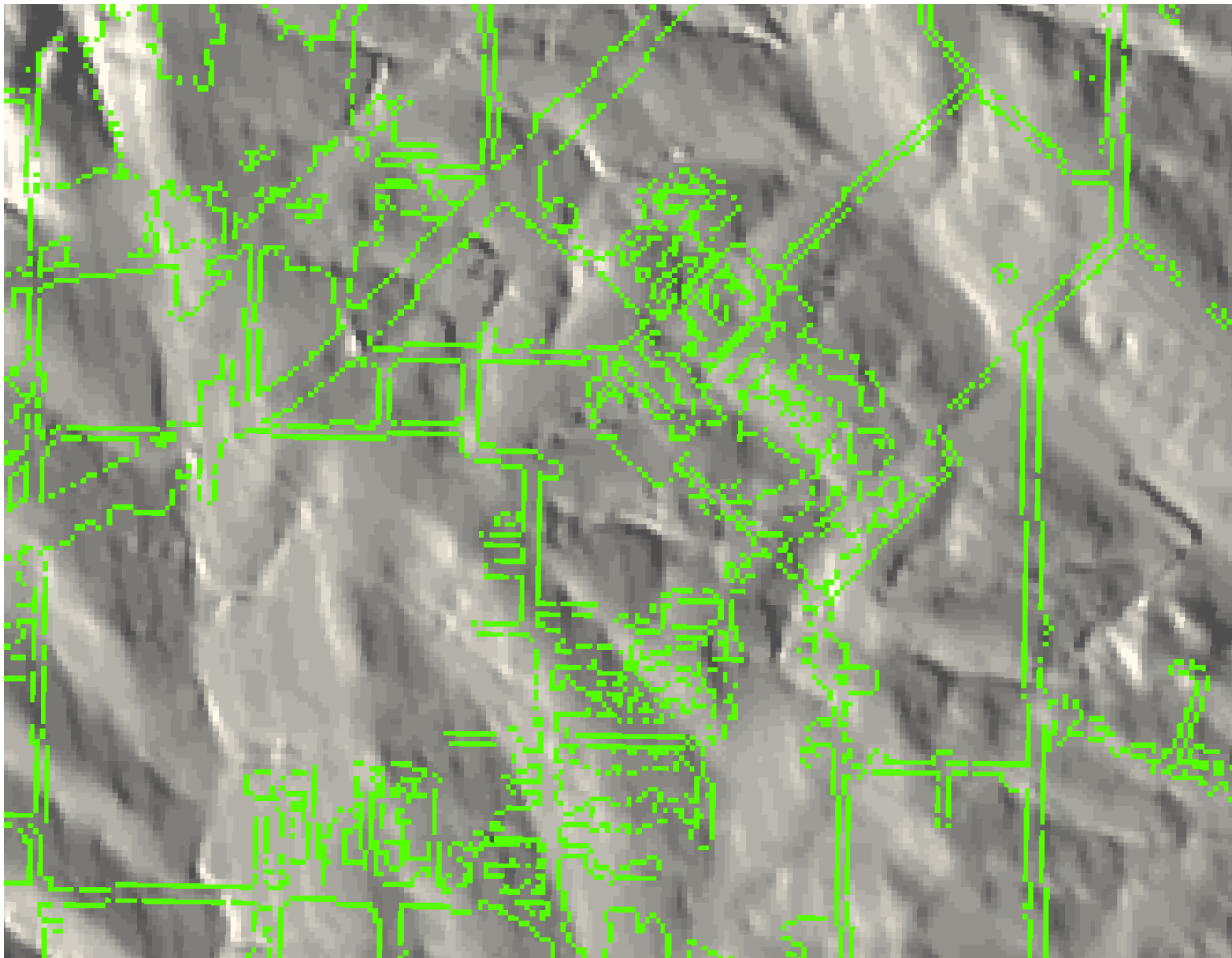




# Urban Landcover and SLEUTH Probable Growth Areas 30m cells



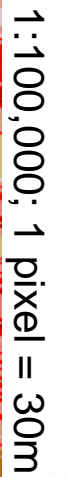
# SLEUTH Probable Growth Areas 30m cells



1:20,000; 1 pixel = 30m

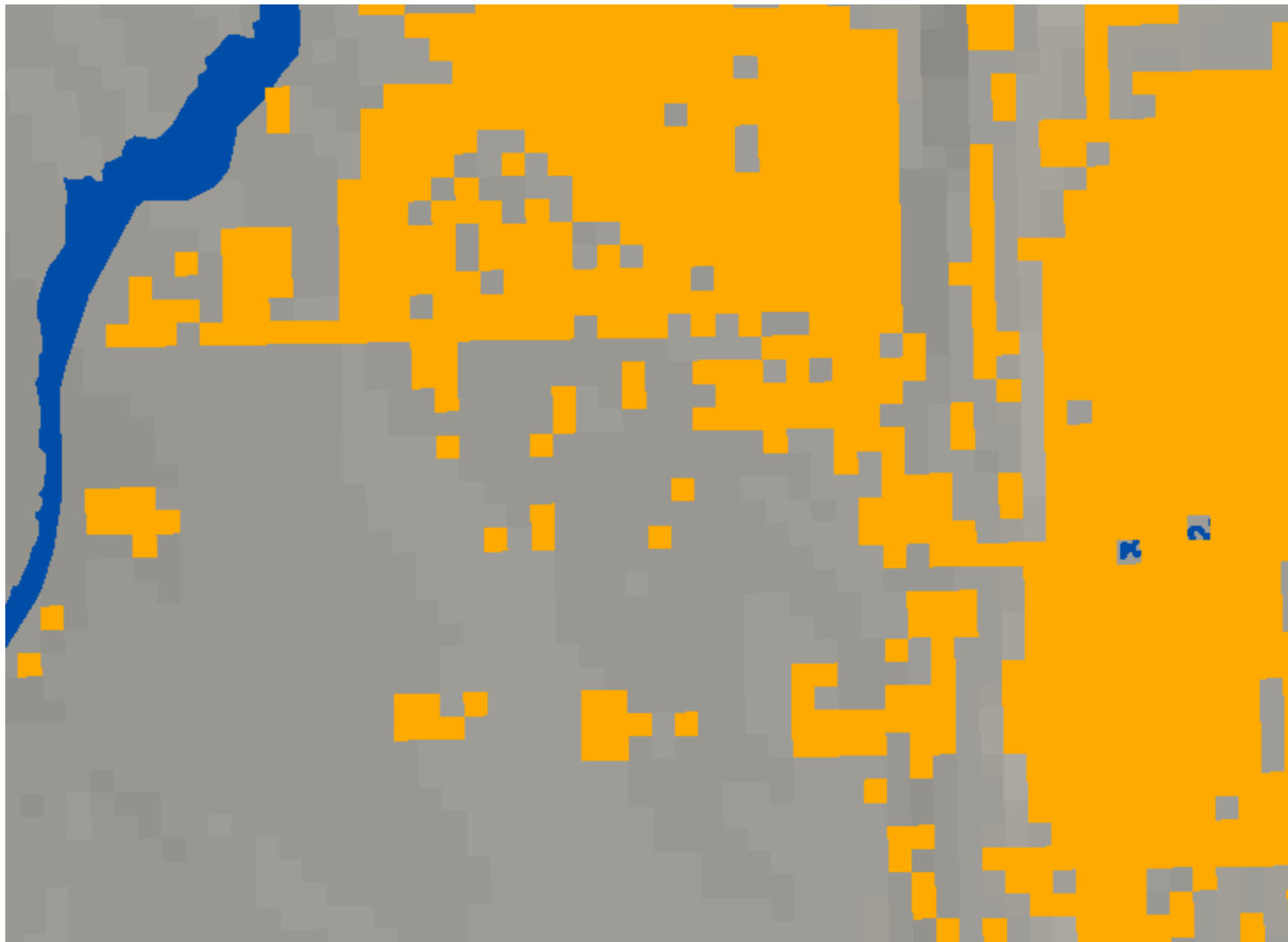


# 2001 NLCD





# 2001 Urban Landcover 360m cells



1:100,000; 1 pixel = 360m



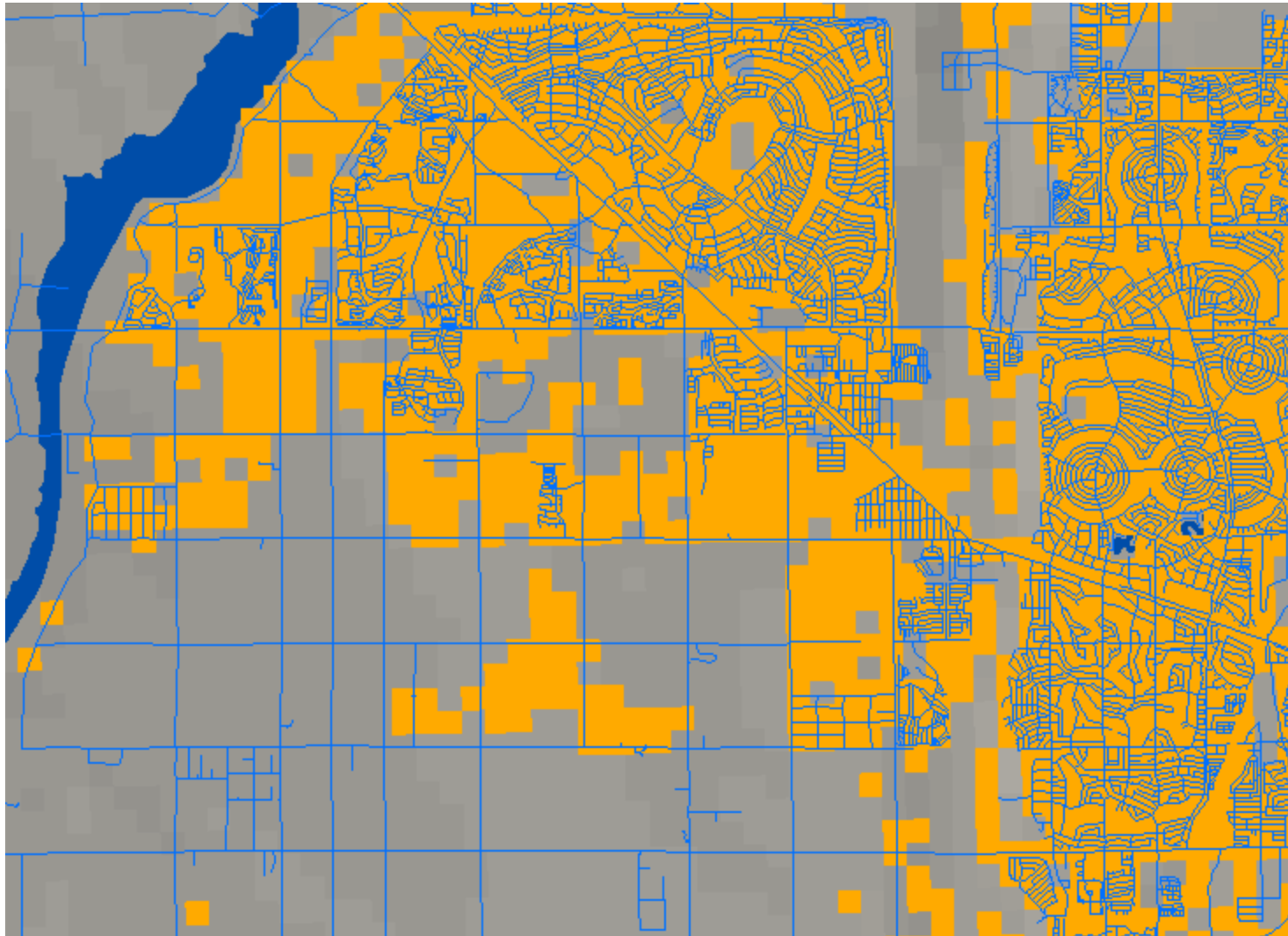
# 2001 Urban Landcover 2001 TIGER Roads



1:100,000; 1 pixel = 360m



# 2001 TIGER Roads 2006 Urban Landcover



1:100,000; 1 pixel = 360m





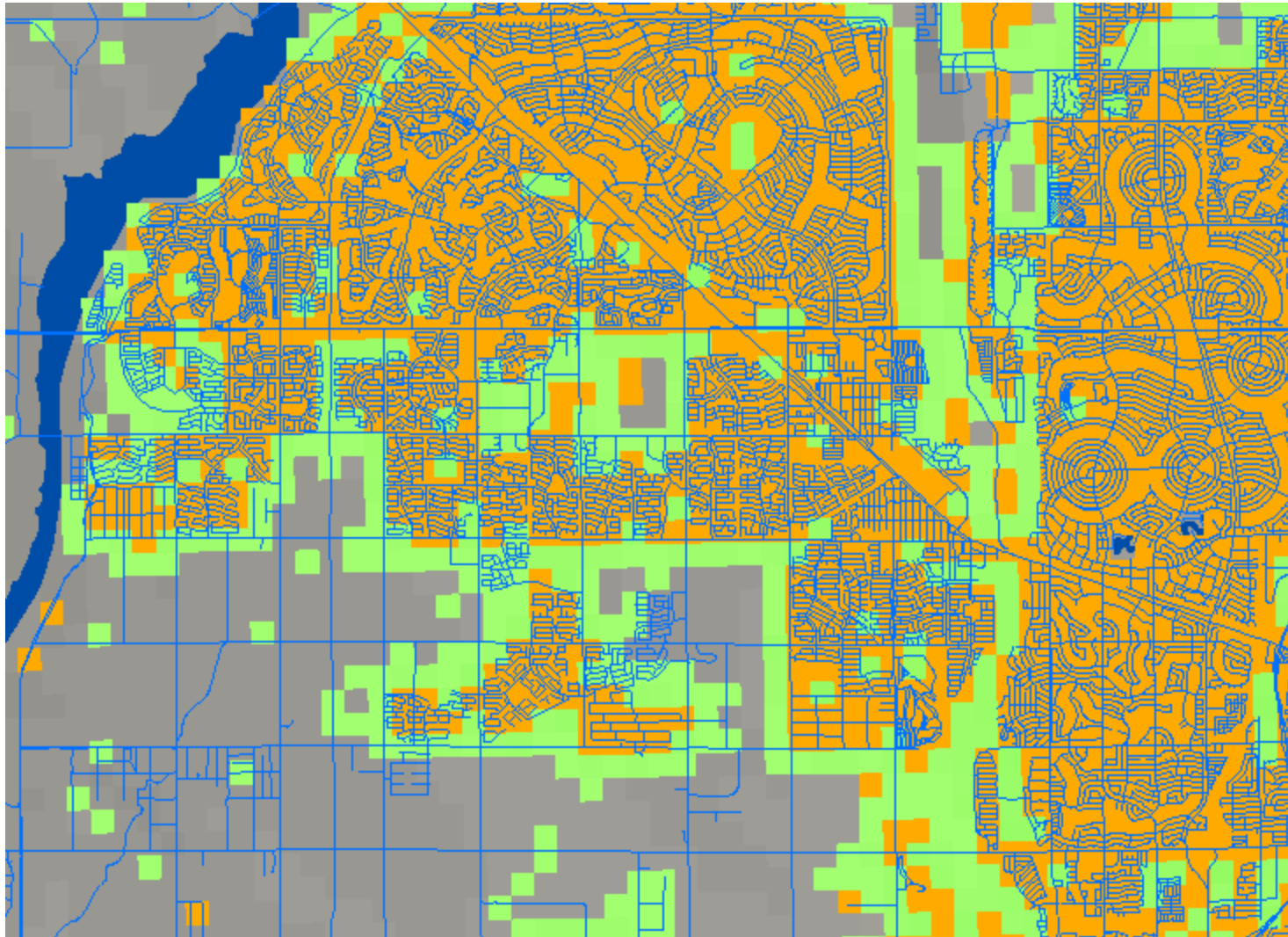
# 2006 Roads



1:125,000; 1 pixel = 360m



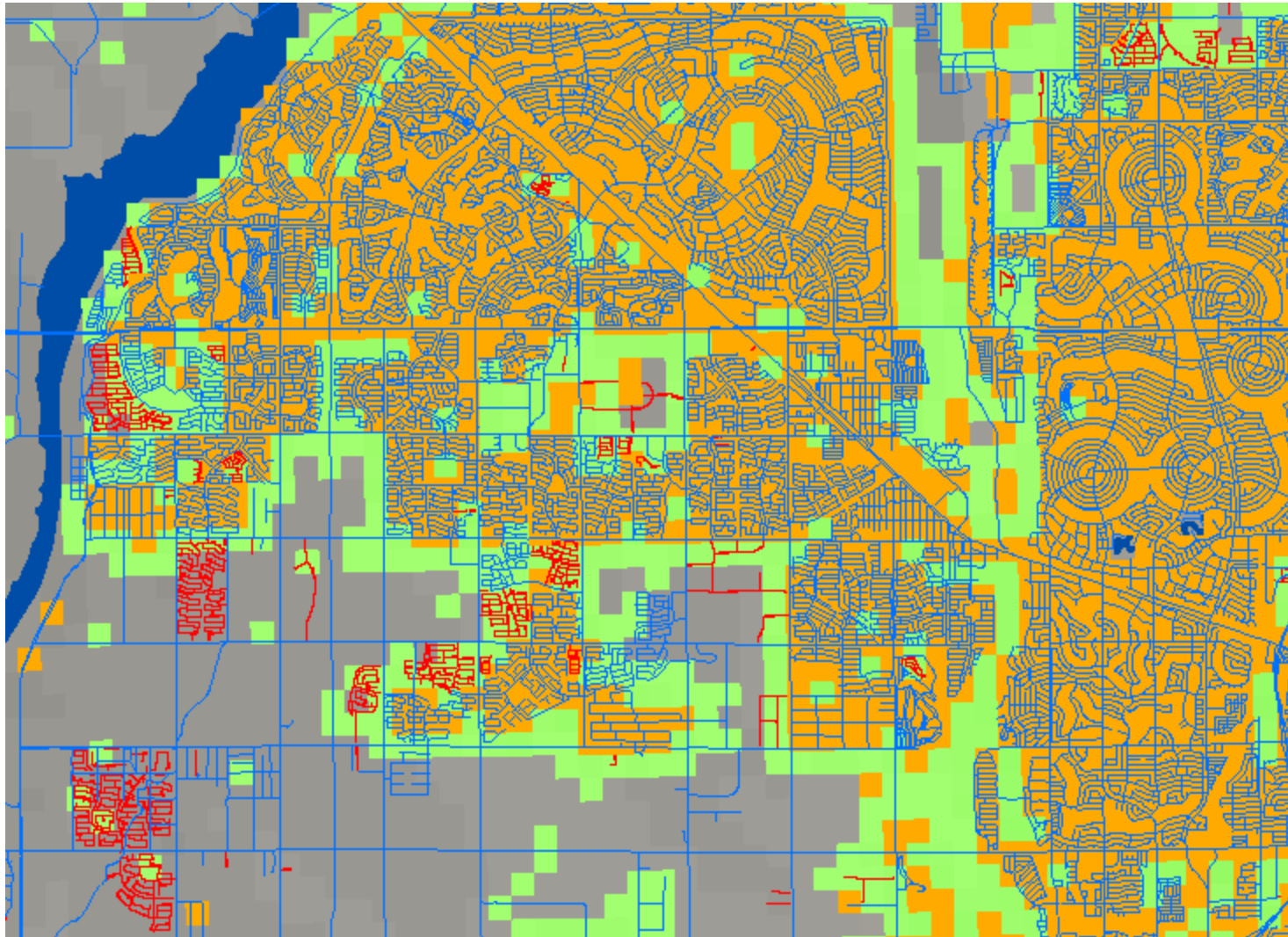
# SLEUTH Probable Growth Areas (GREEN) 360m cells



1:100,000; 1 pixel = 360m



# SLEUTH Probable Growth Areas and Actual Road Growth to 2011 (Red)

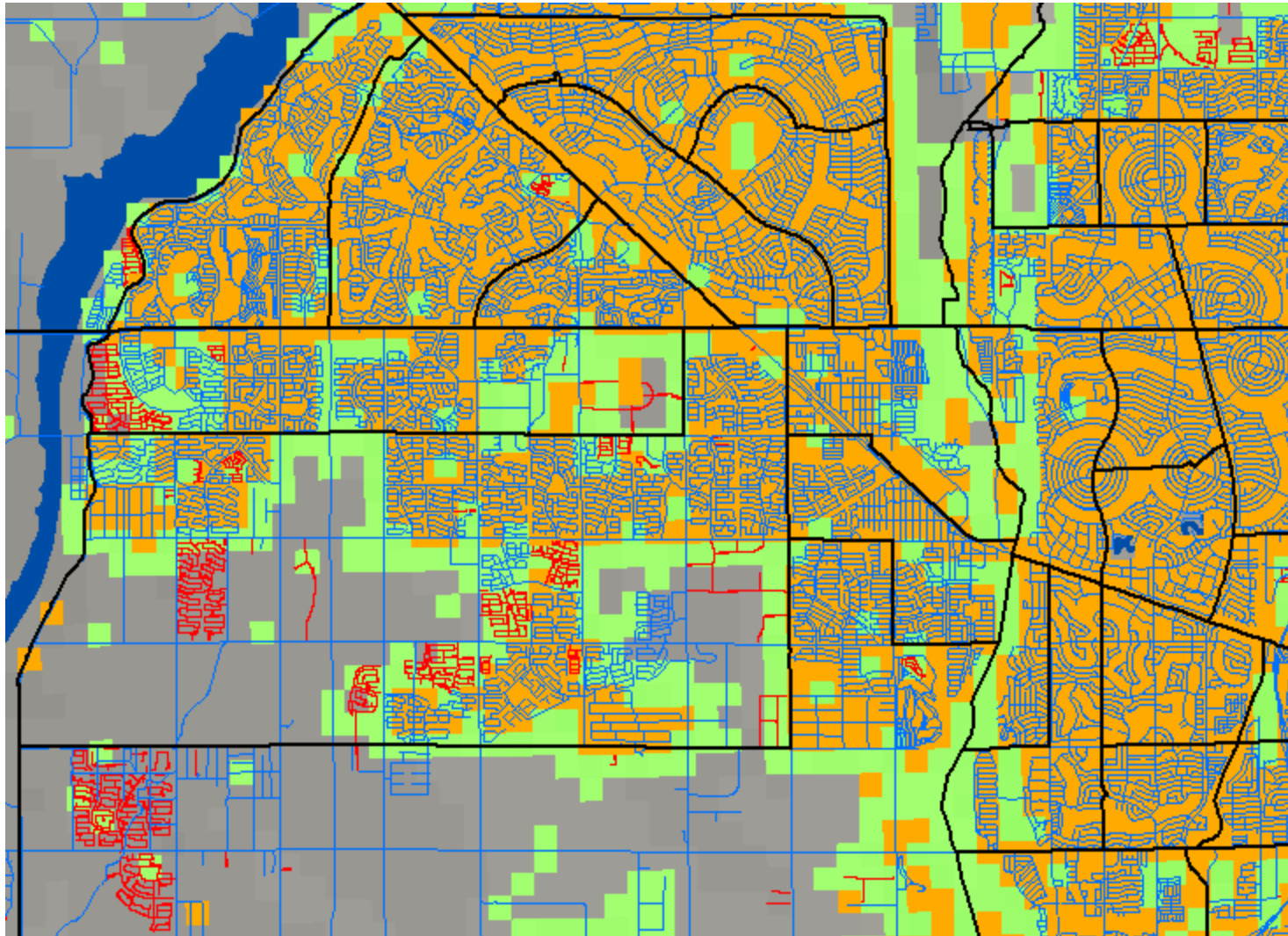


1:100,000; 1 pixel = 360m





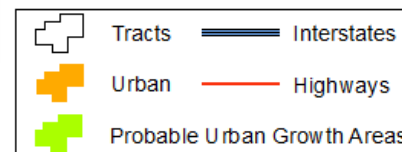
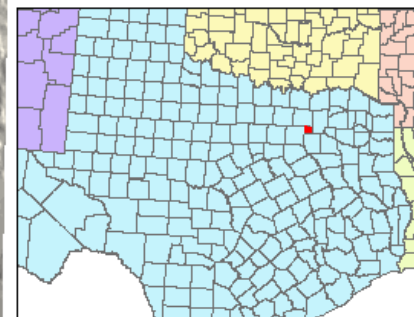
# SLEUTH Probable Growth Areas and Actual Road Growth to 2011 with Tract Boundaries



1:100,000; 1 pixel = 360m



# Study Area 1: Rockwall County, TX

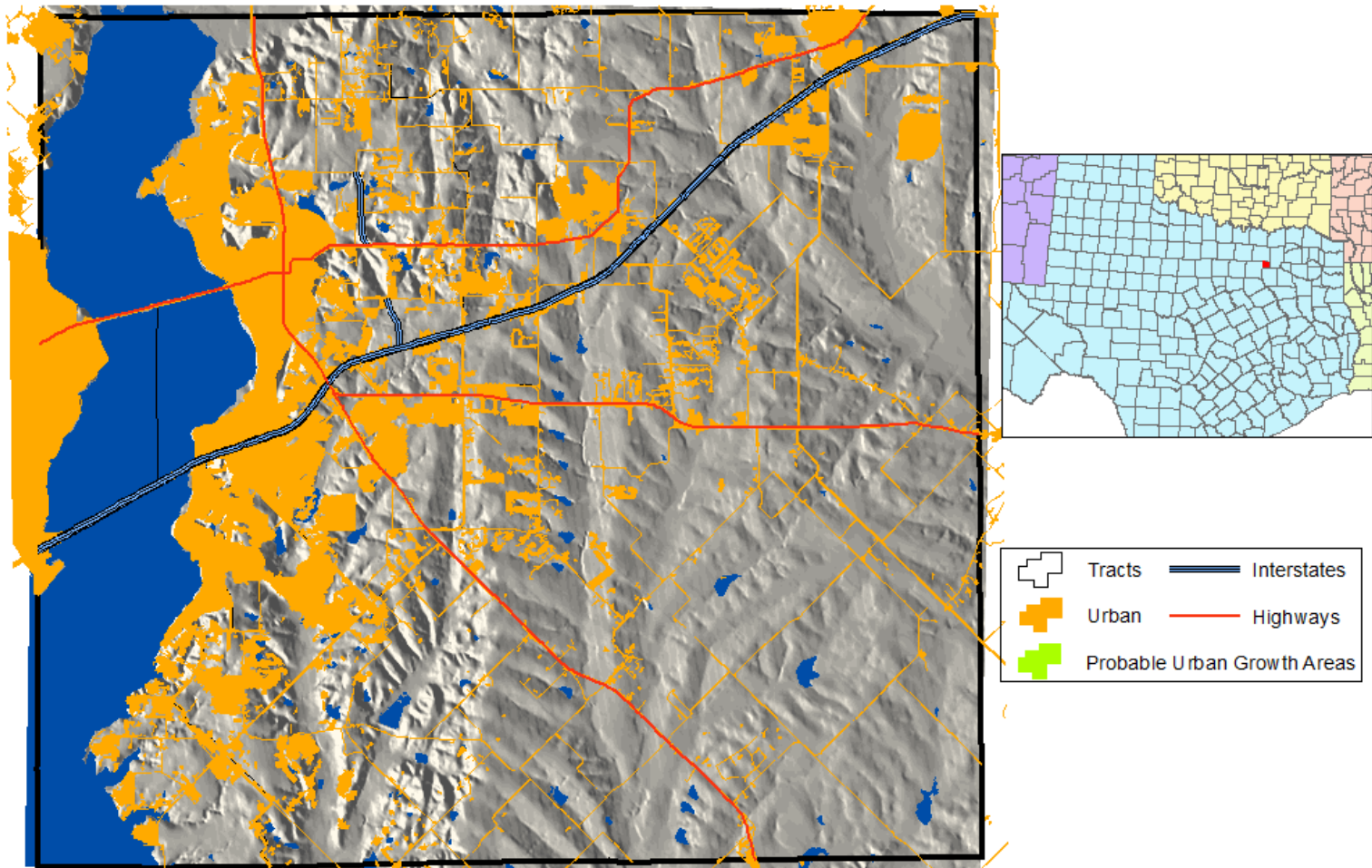


149 Sq. Mi.

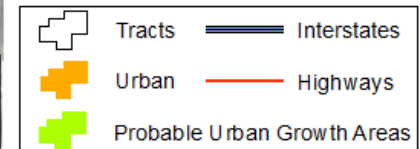
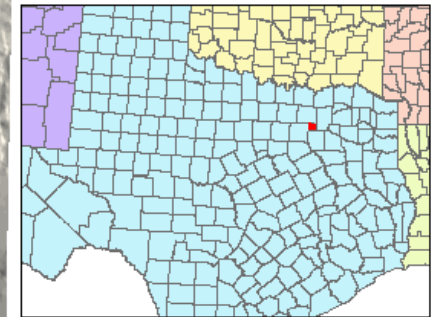




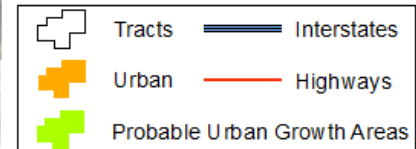
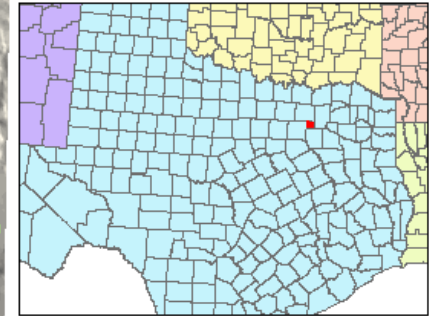
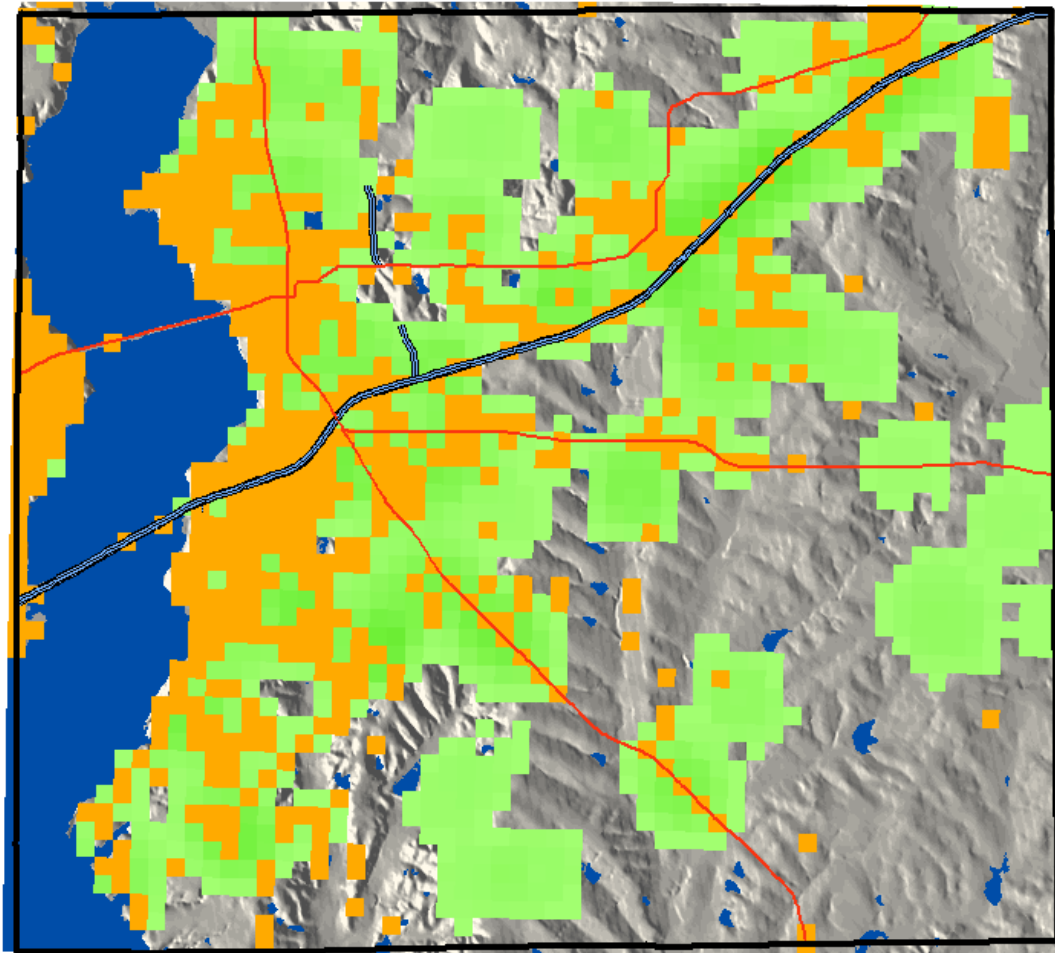
## Study Area 1: Rockwall County, TX







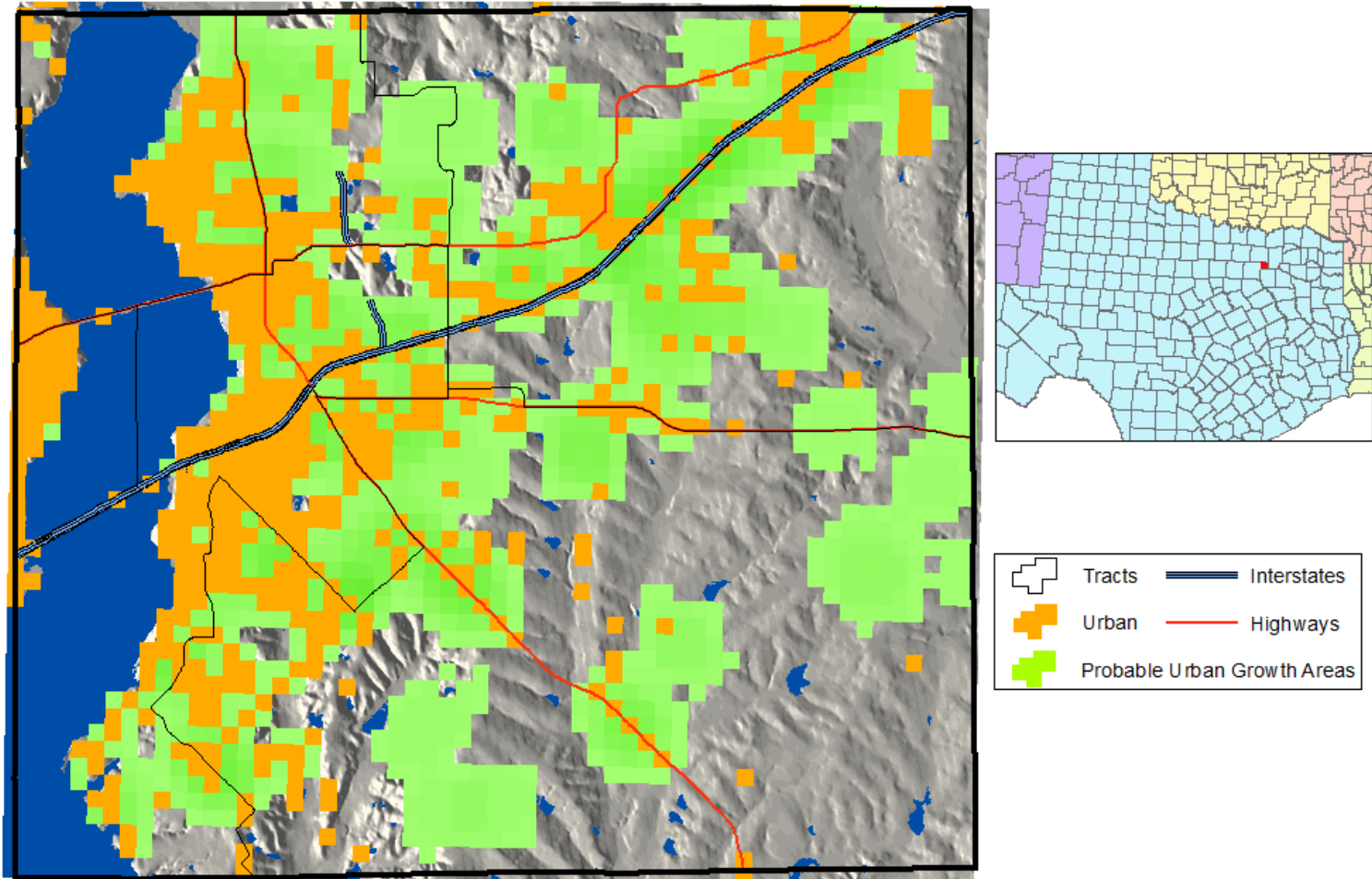
# Study Area 1: Rockwall County, TX



SLEUTH Probable Growth Areas (2006-2011)



# Study Area 1: Rockwall County, TX

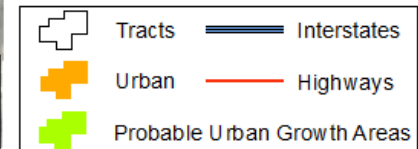
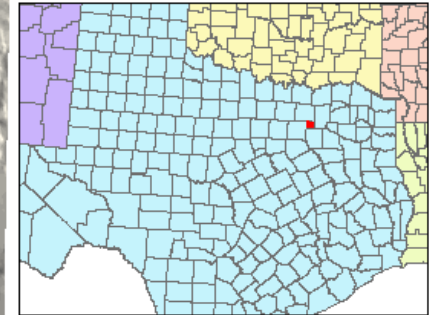
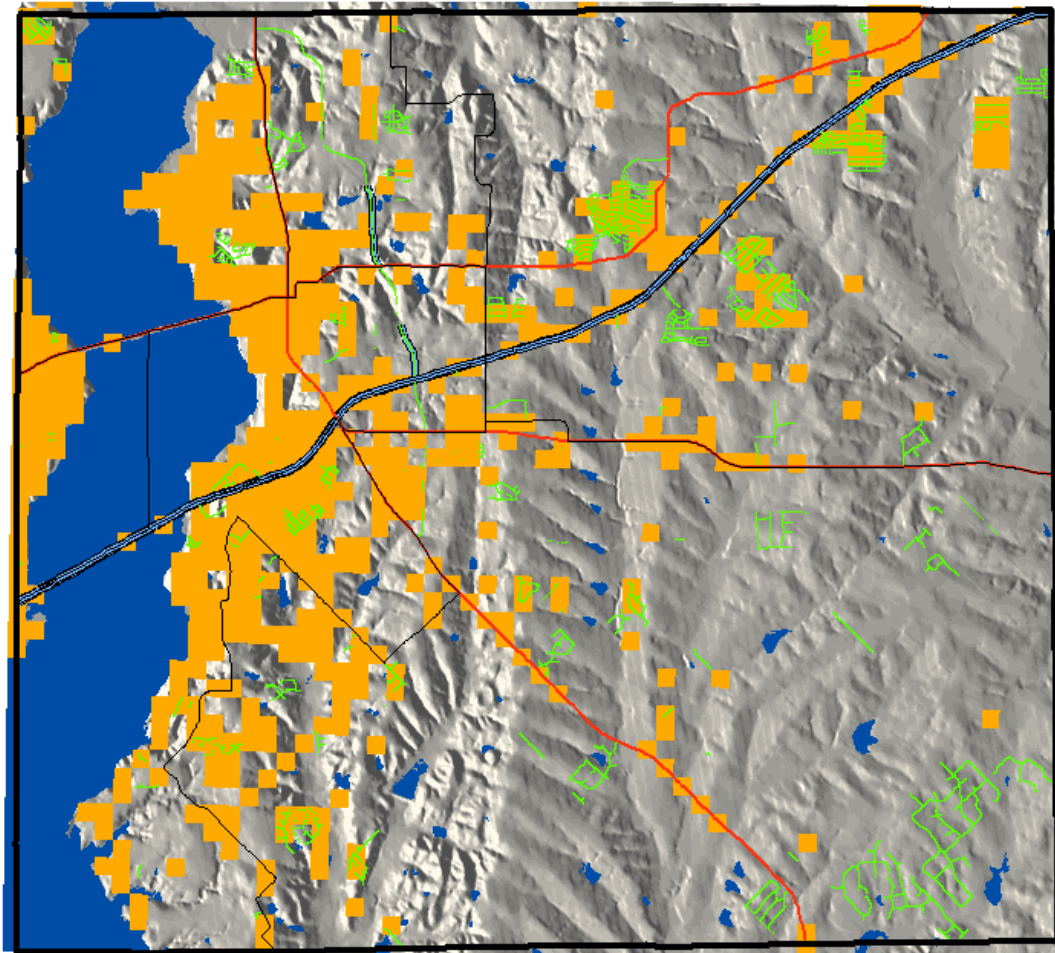


SLEUTH Probable Growth Areas (2006-2011) and Tracts (black)



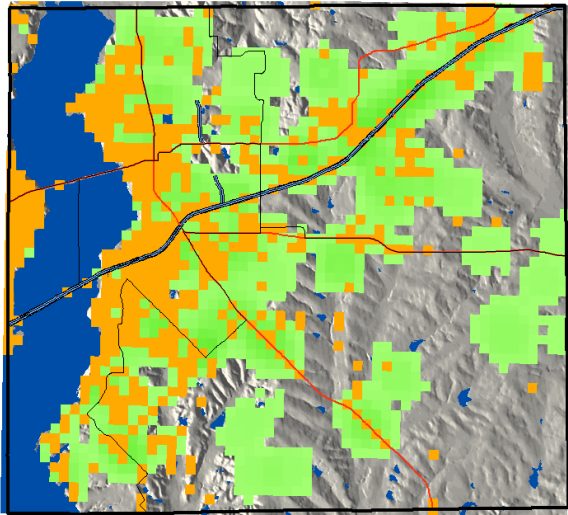


# Study Area 1: Rockwall County, TX

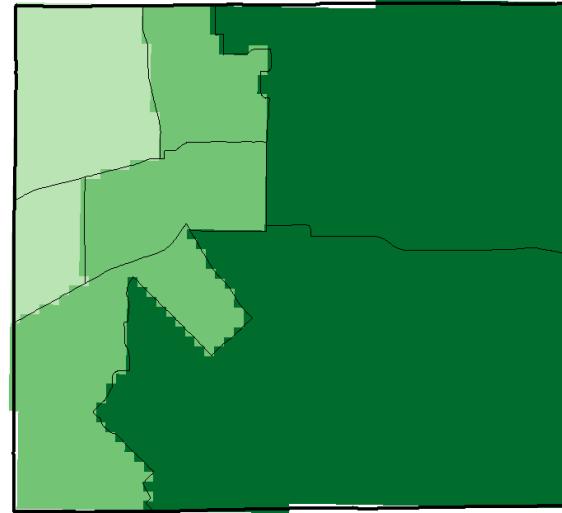


Actual Road Growth (2006-2011) and Tracts (black)



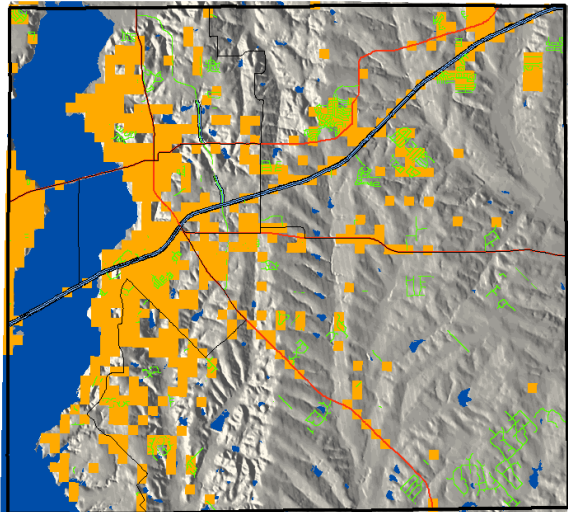
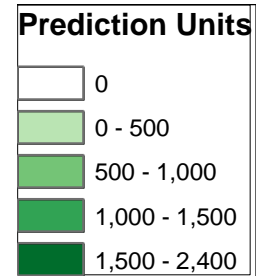


SLEUTH Probable Growth Areas

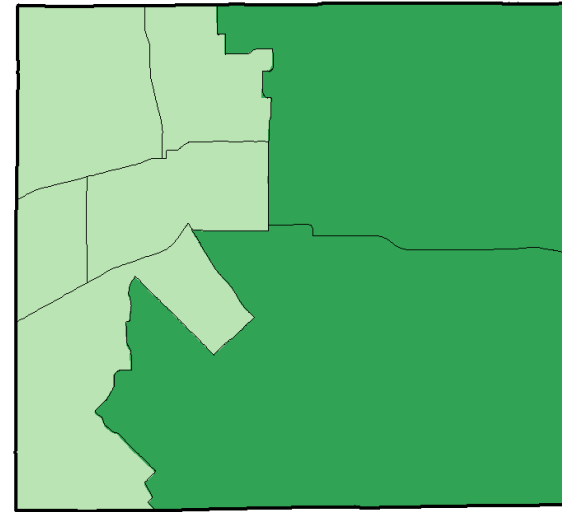


SLEUTH Summed by Tract

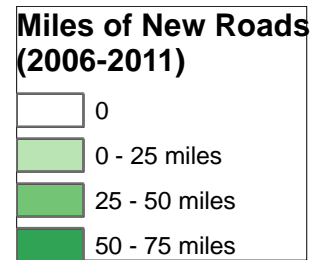
(Summed % Probability of New Growth for all 360m Cells)



Actual Road Growth

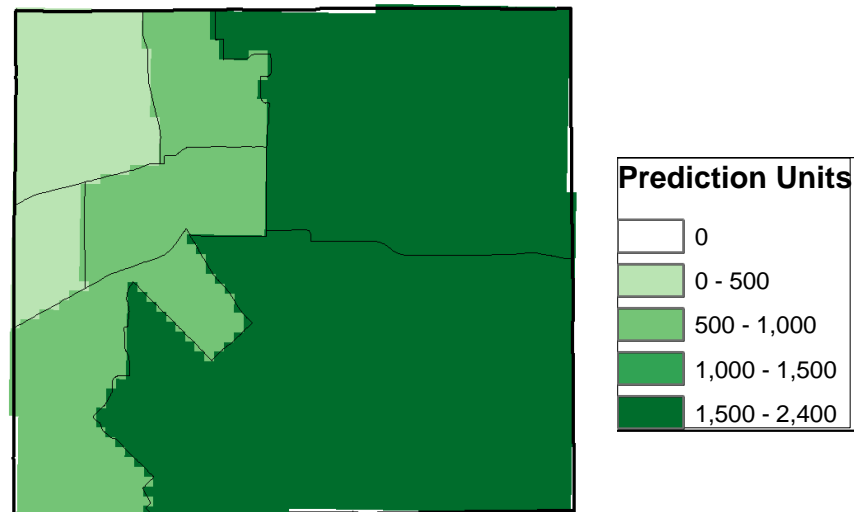


Road Growth Summed by Tract

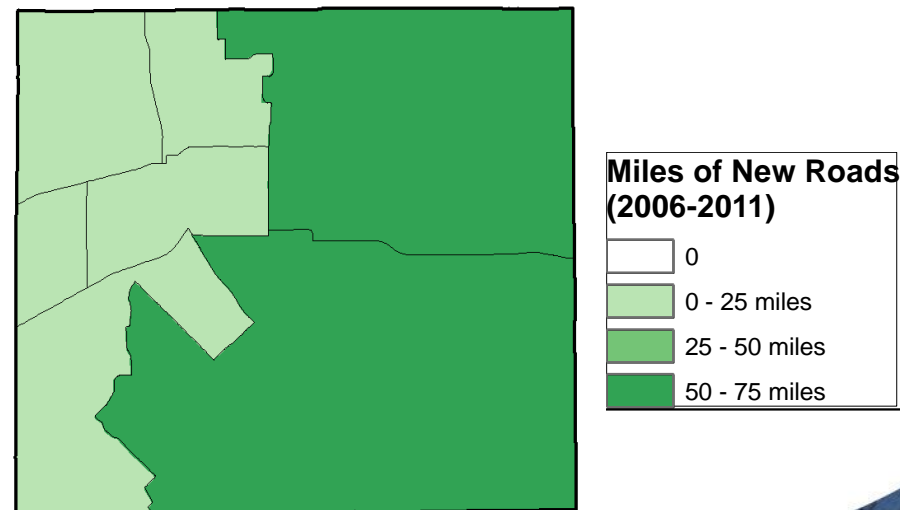


Rockwall, TX Tract ID	SLEUTH Prediction Units ("Cumulate Urban")	Actual Road Growth (miles)
48397040100	229	8.3
48397040200	792	11.1
48397040301	2	0.4
48397040302	622	5.3
48397040400	3,445	54.3
48397040501	3,189	58.8
48397040502	872	7.2

Correlation = 0.98



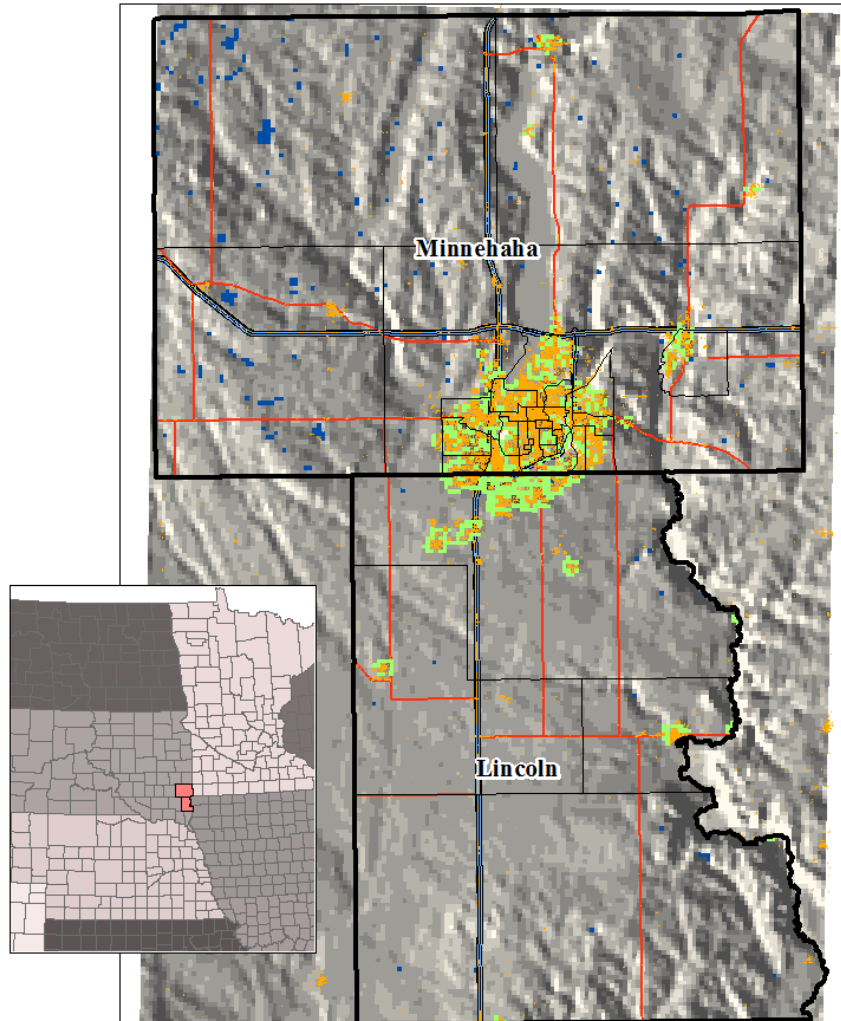
**SLEUTH Summed by Tract**  
(Summed % Probability of New Growth for all 360m Cells)



**Road Growth Summed by Tract**



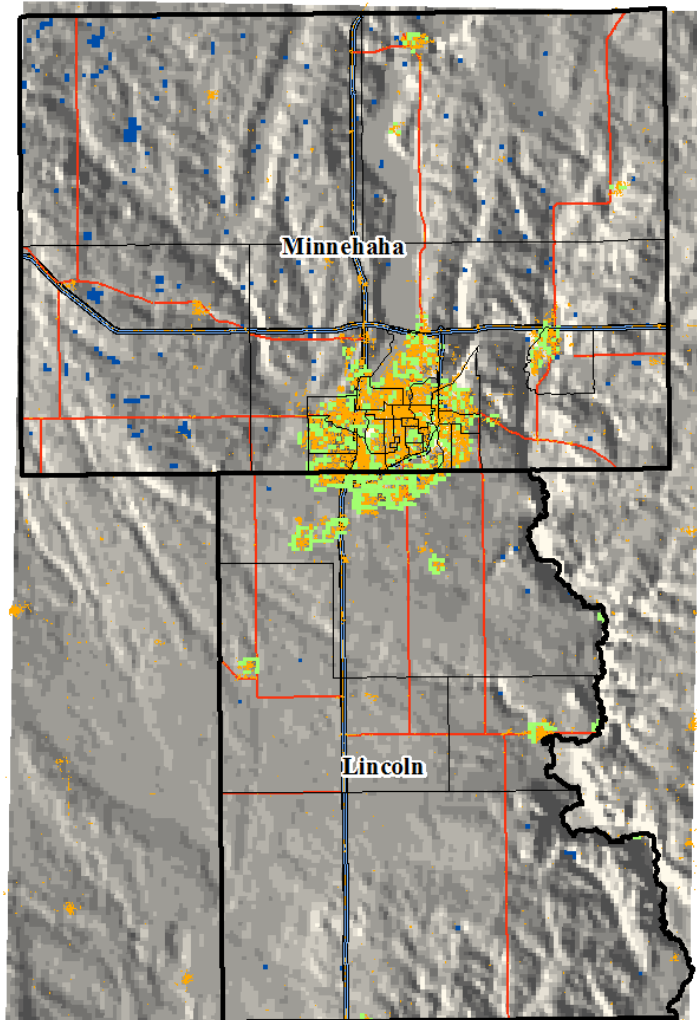
# Study Area 2: Minnehaha and Lincoln Counties, SD



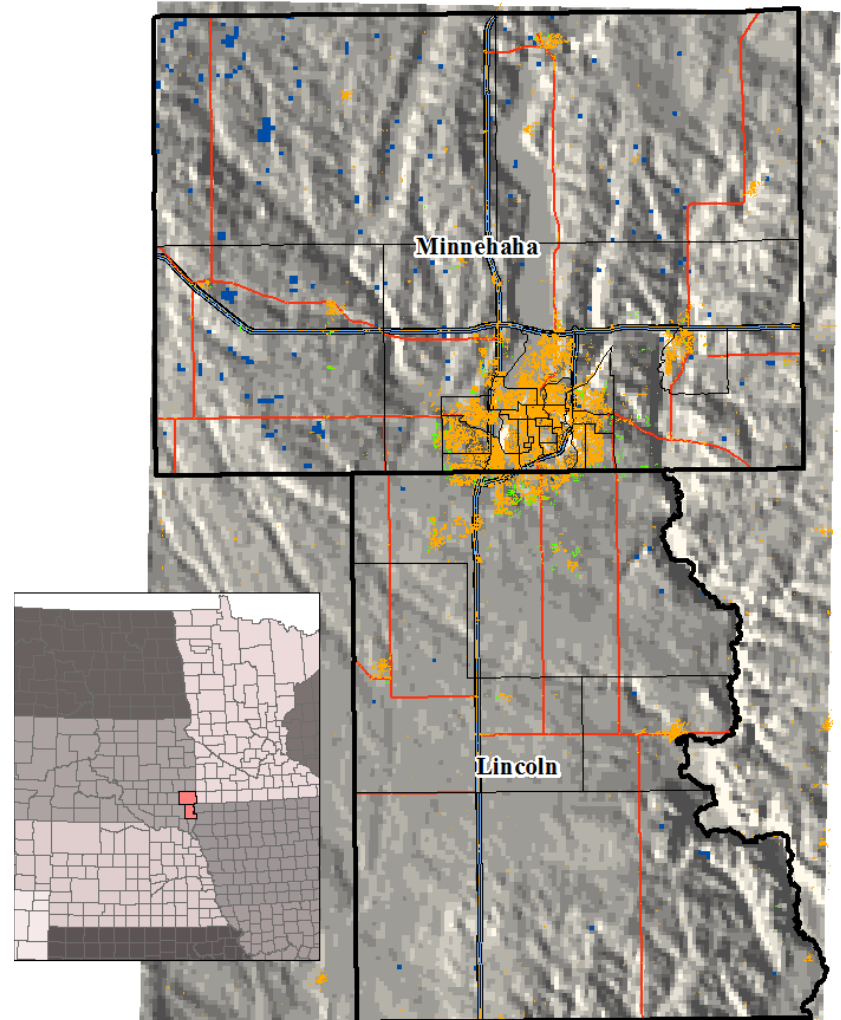
1,387 Sq. Mi.



# Study Area 2: Minnehaha and Lincoln Counties, SD



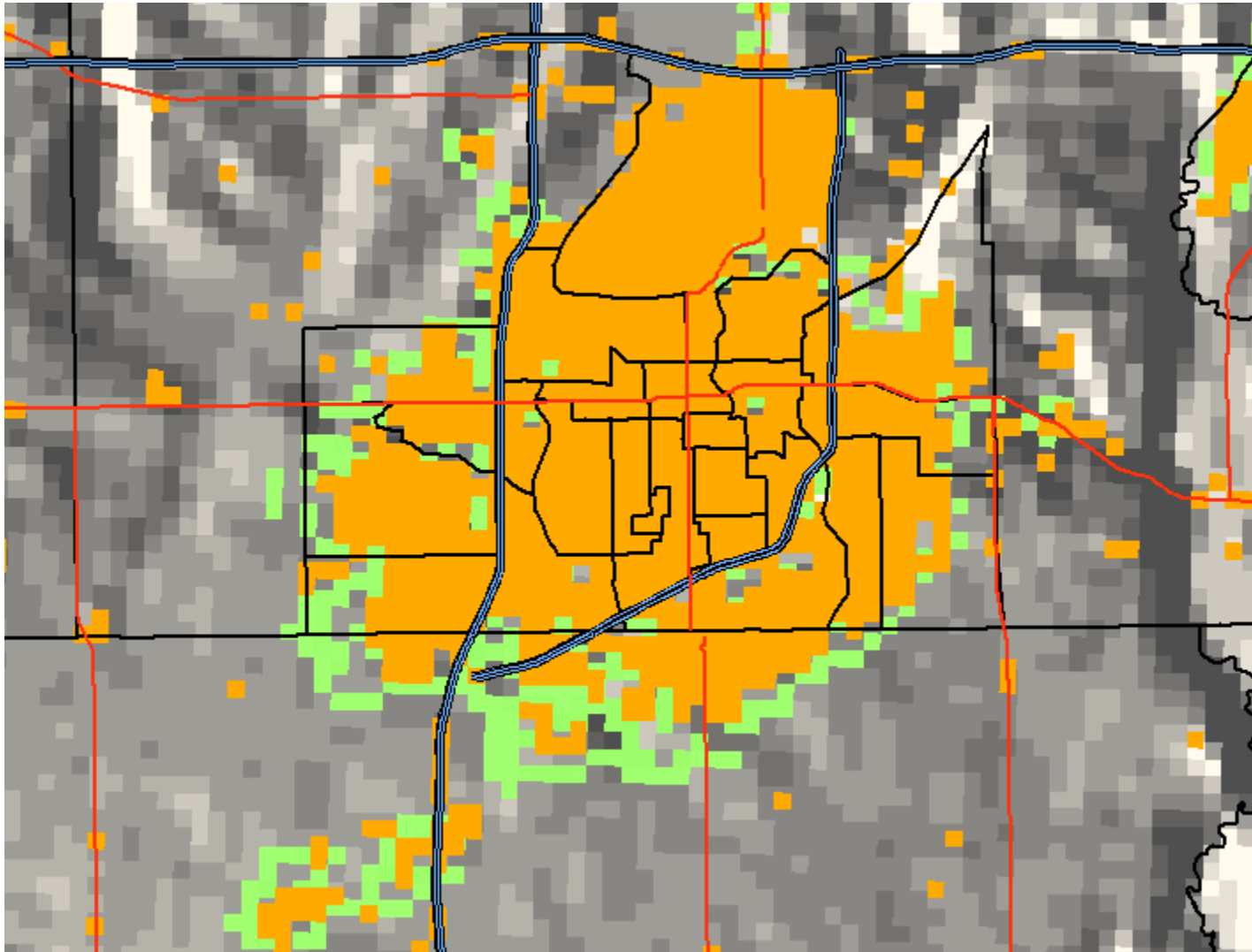
SLEUTH Probable Growth Areas



Actual Road Growth

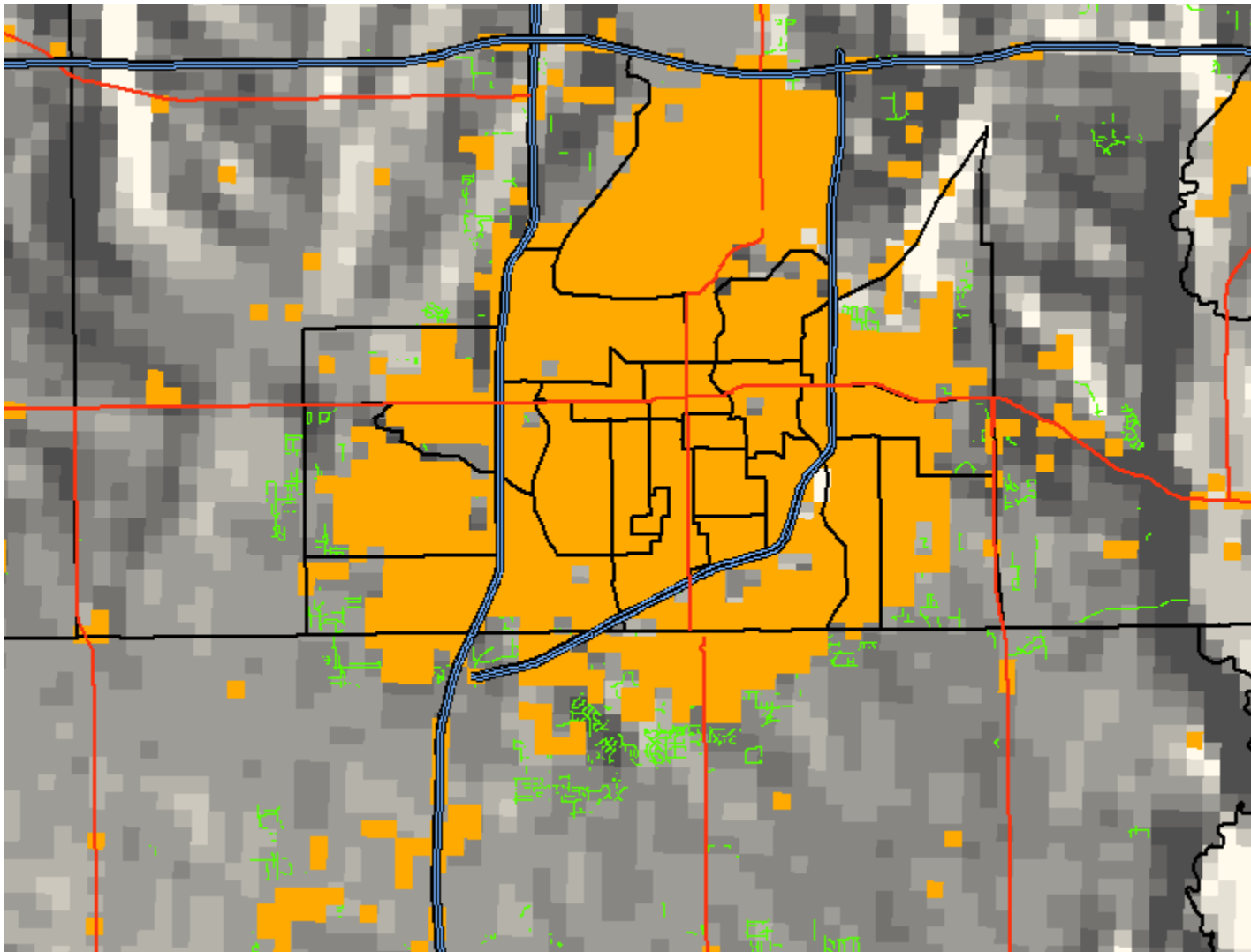


## Study Area 2: Minnehaha and Lincoln Counties, SD

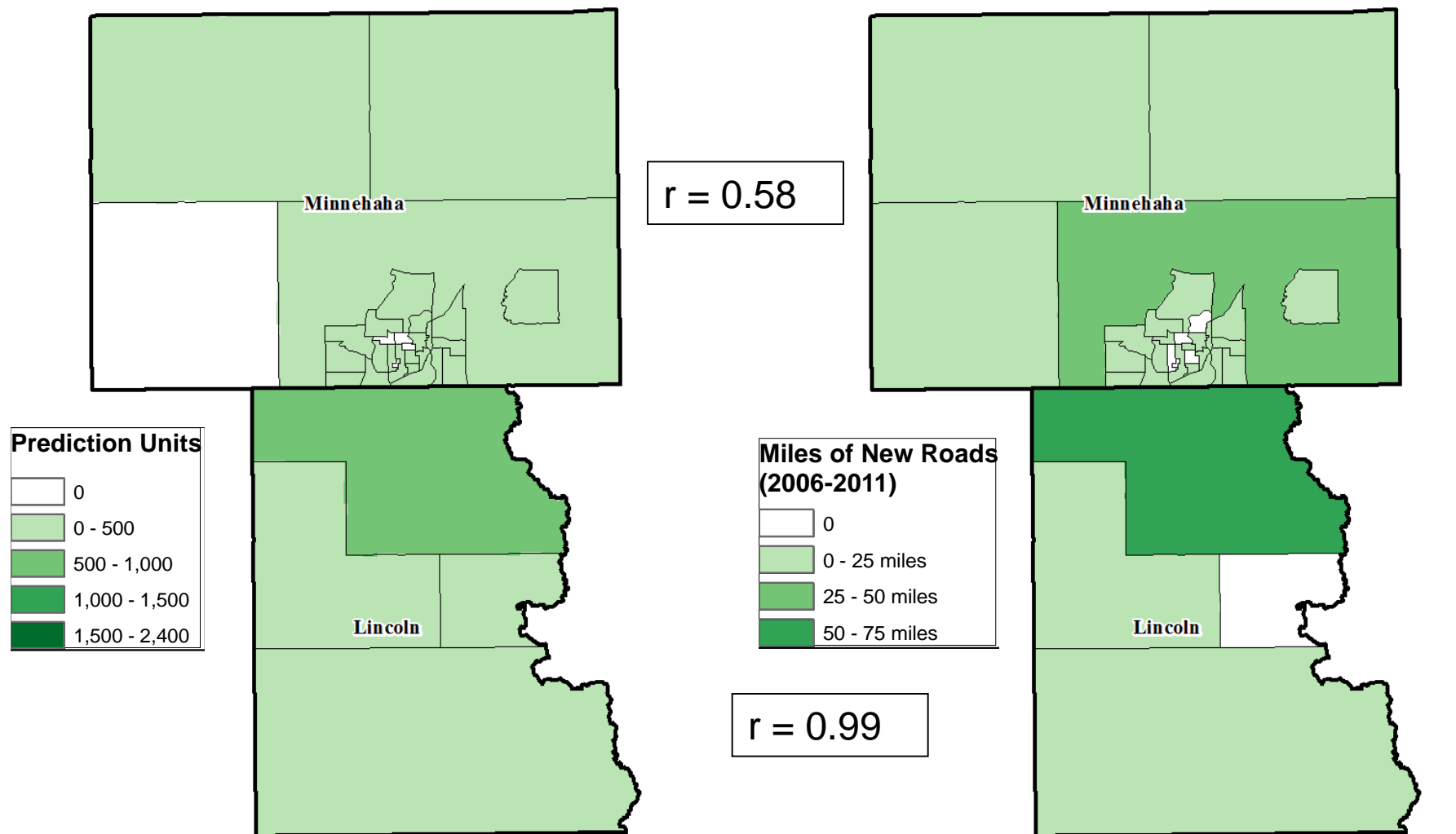




## Study Area 2: Minnehaha and Lincoln Counties, SD



# Study Area 2: Minnehaha and Lincoln Counties, SD



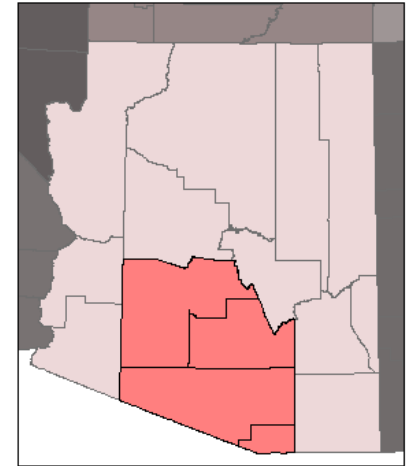
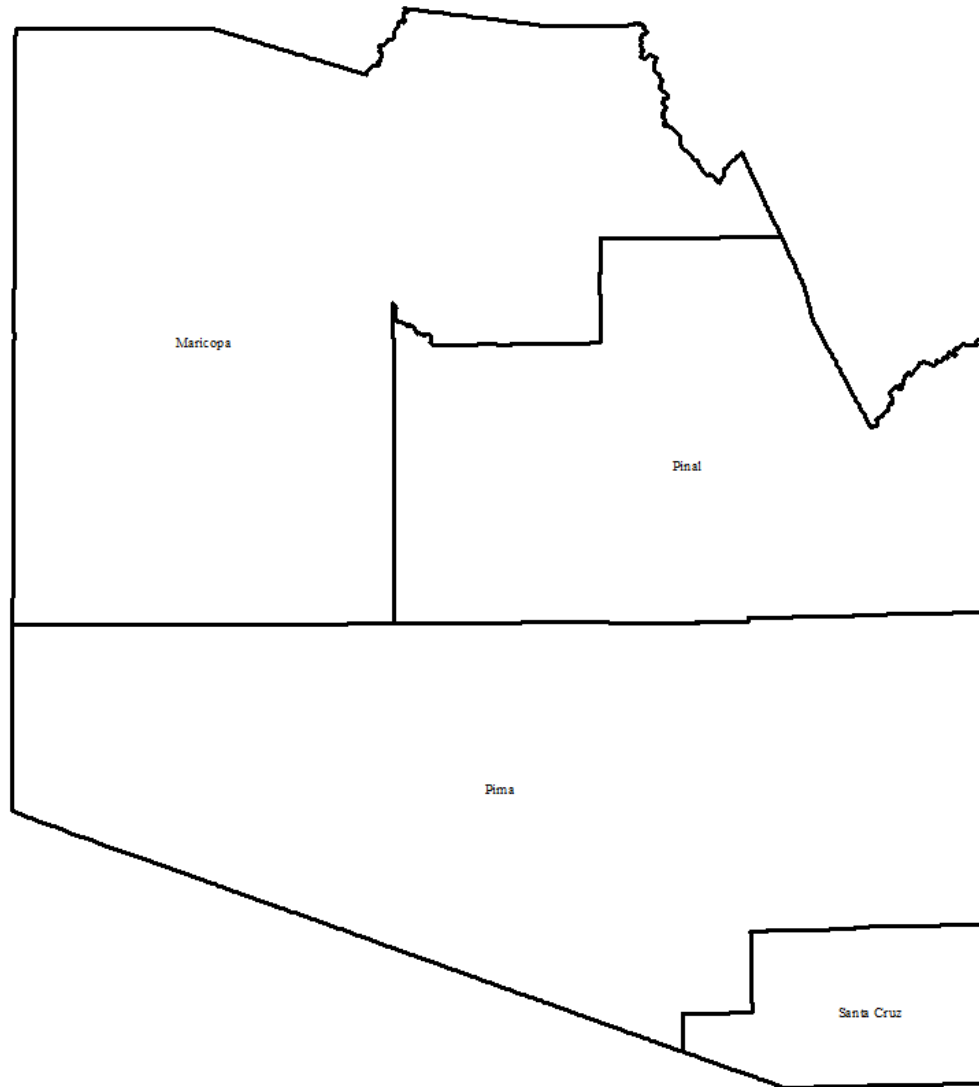
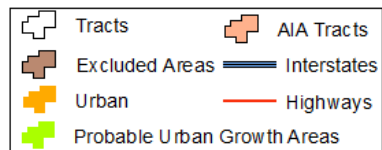
**SLEUTH Summed by Tract**

(Summed % Probability of New Growth for all 360m Cells)

**Road Growth Summed by Tract**



## Study Area 3: Arizona Counties



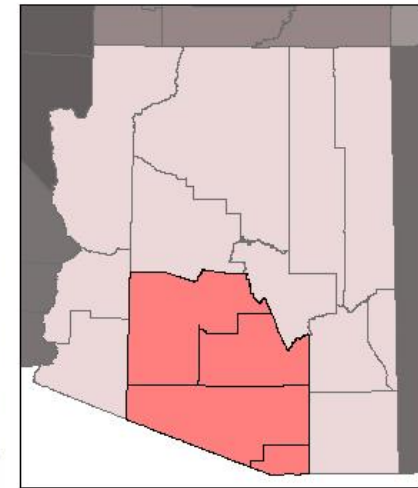
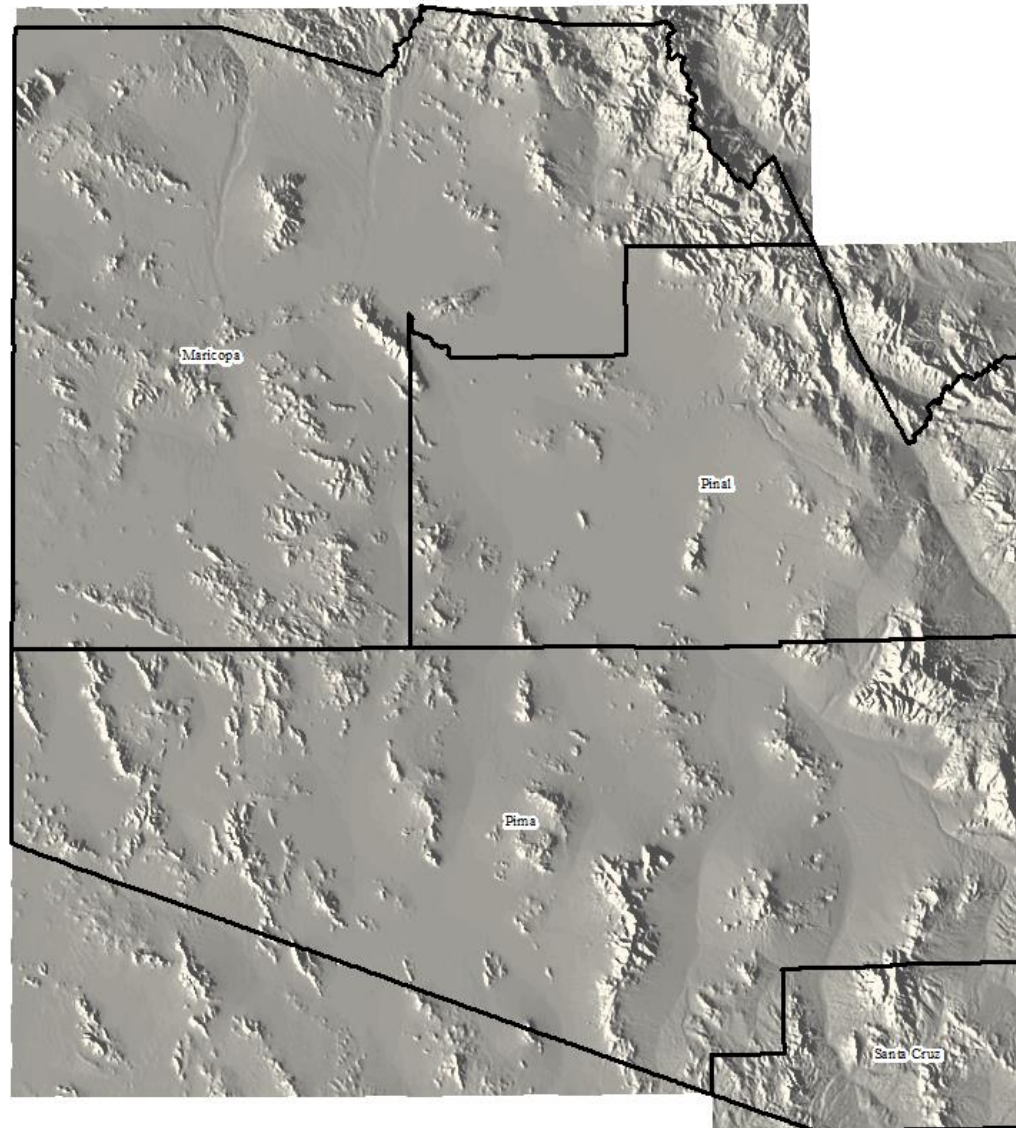
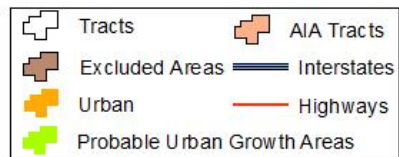
25,000 Sq. Mi.



## “URBAN DNA”

COUNTY	Diffusion	Breed	Spread	Slope Resistance	Road Gravity
Maricopa, AZ	75	50	1	50	1
Pima, AZ	25	50	1	25	75
Pinal, AZ	1	50	1	25	100
Santa Cruz, AZ	1	1	1	75	75
Lincoln, SD	1	1	1	1	75
Minnehaha, SD	1	1	1	1	50
Rockwall, TX	1	100	1	75	25

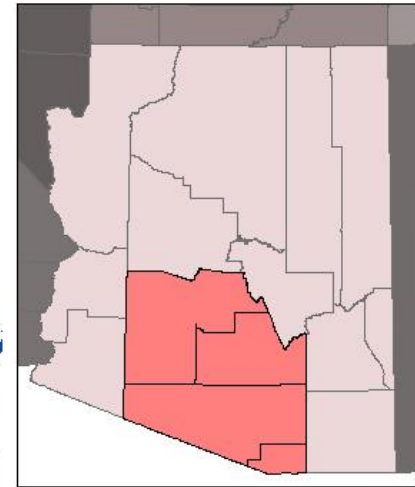
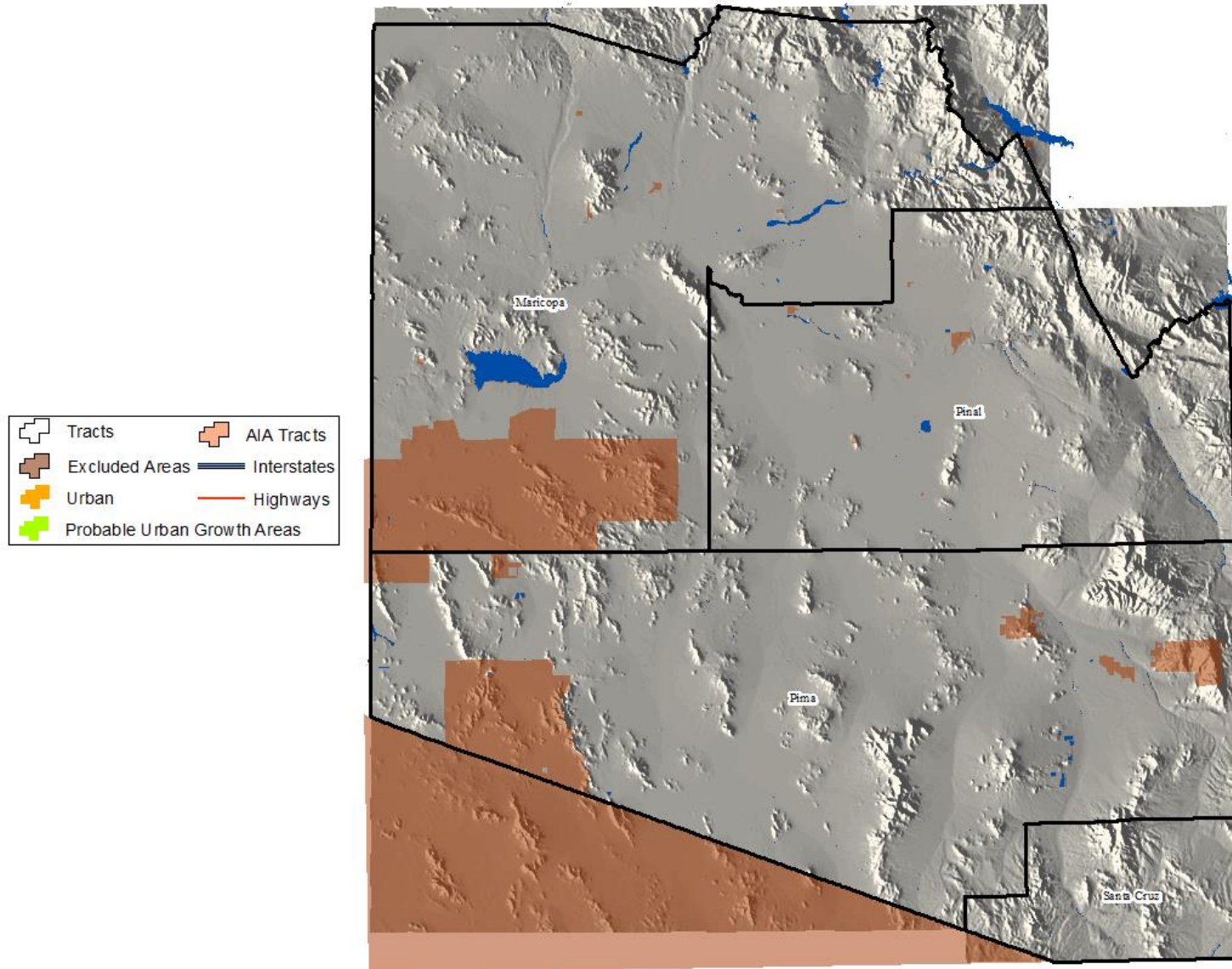
# Study Area 3: Arizona Counties



25,000 Sq. Mi.

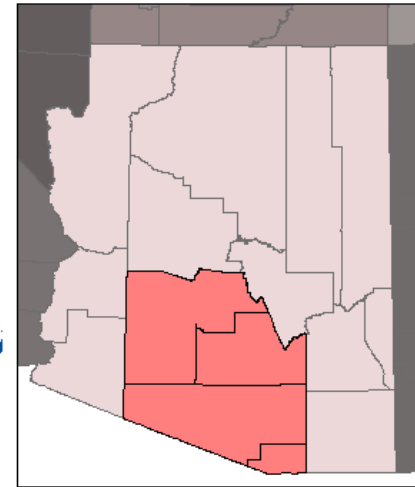
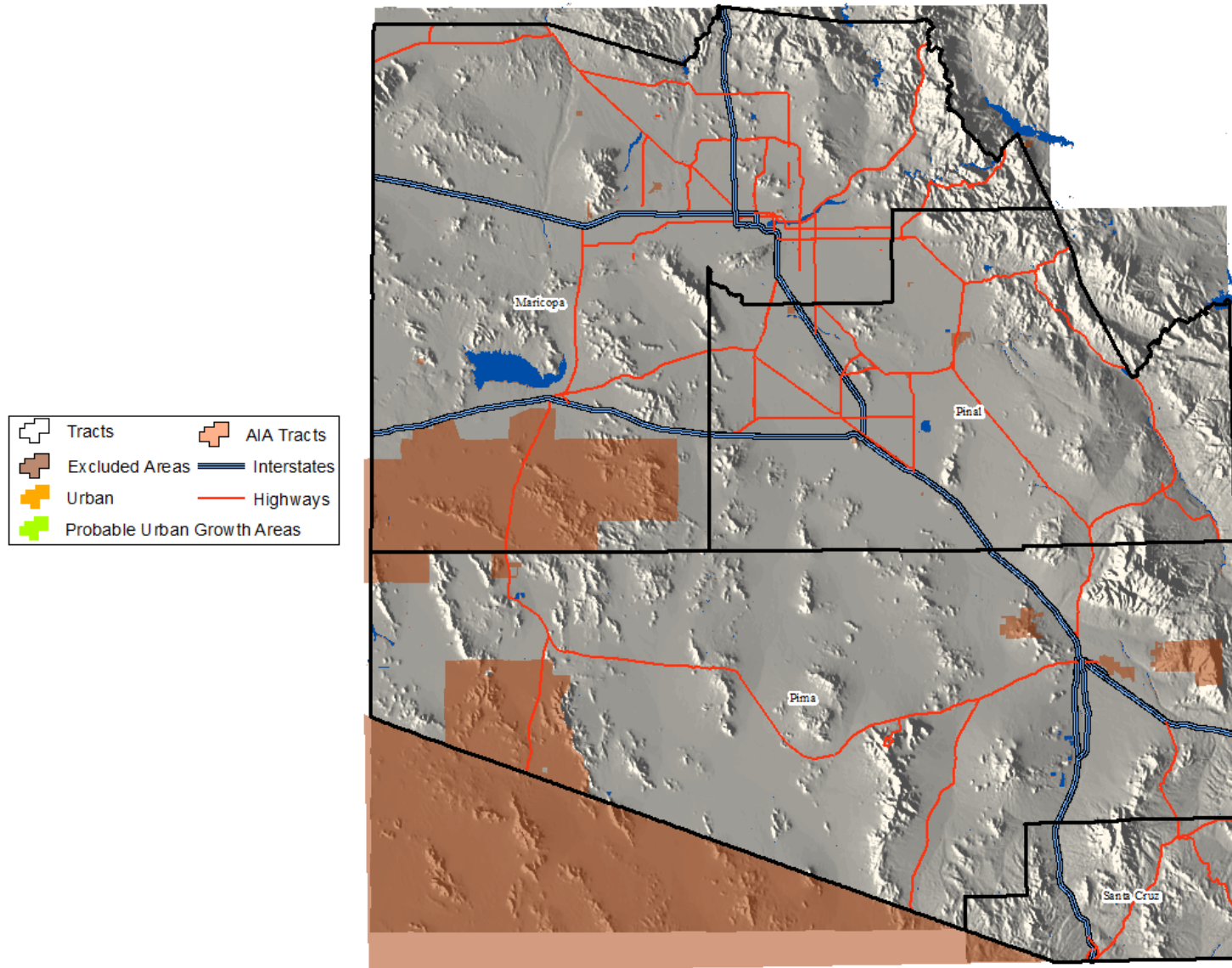


# Study Area 3: Arizona Counties

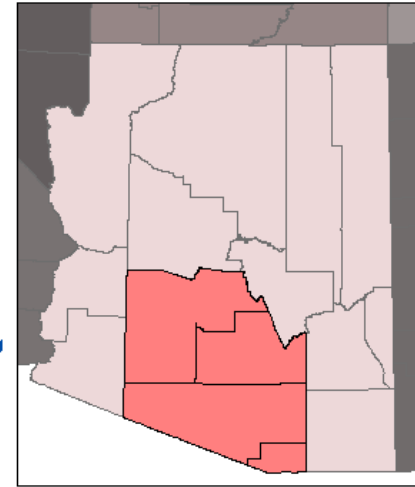
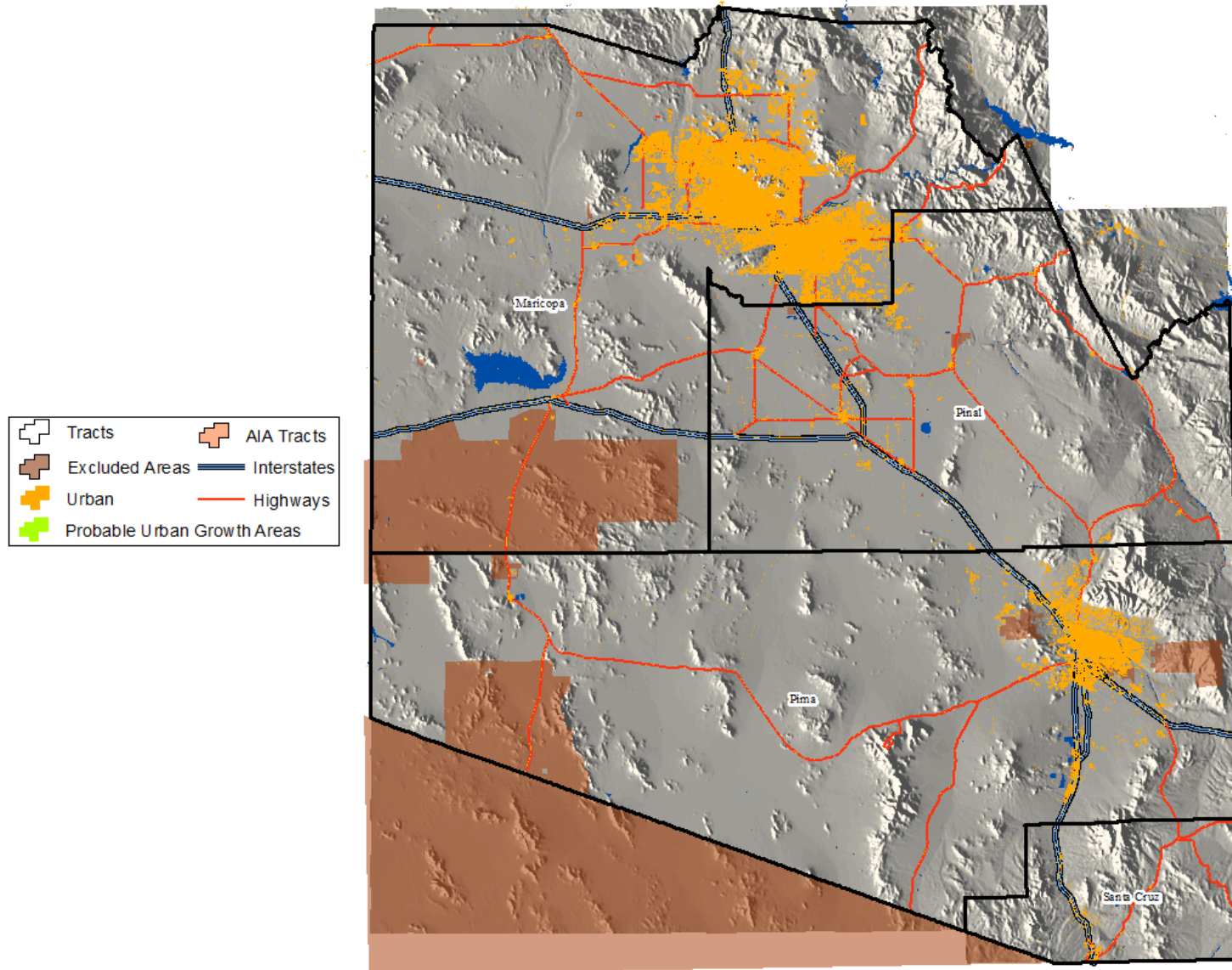




# Study Area 3: Arizona Counties

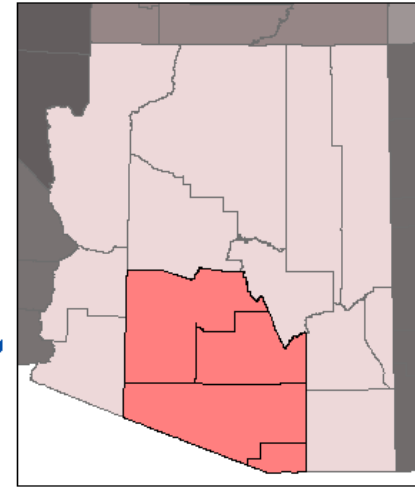
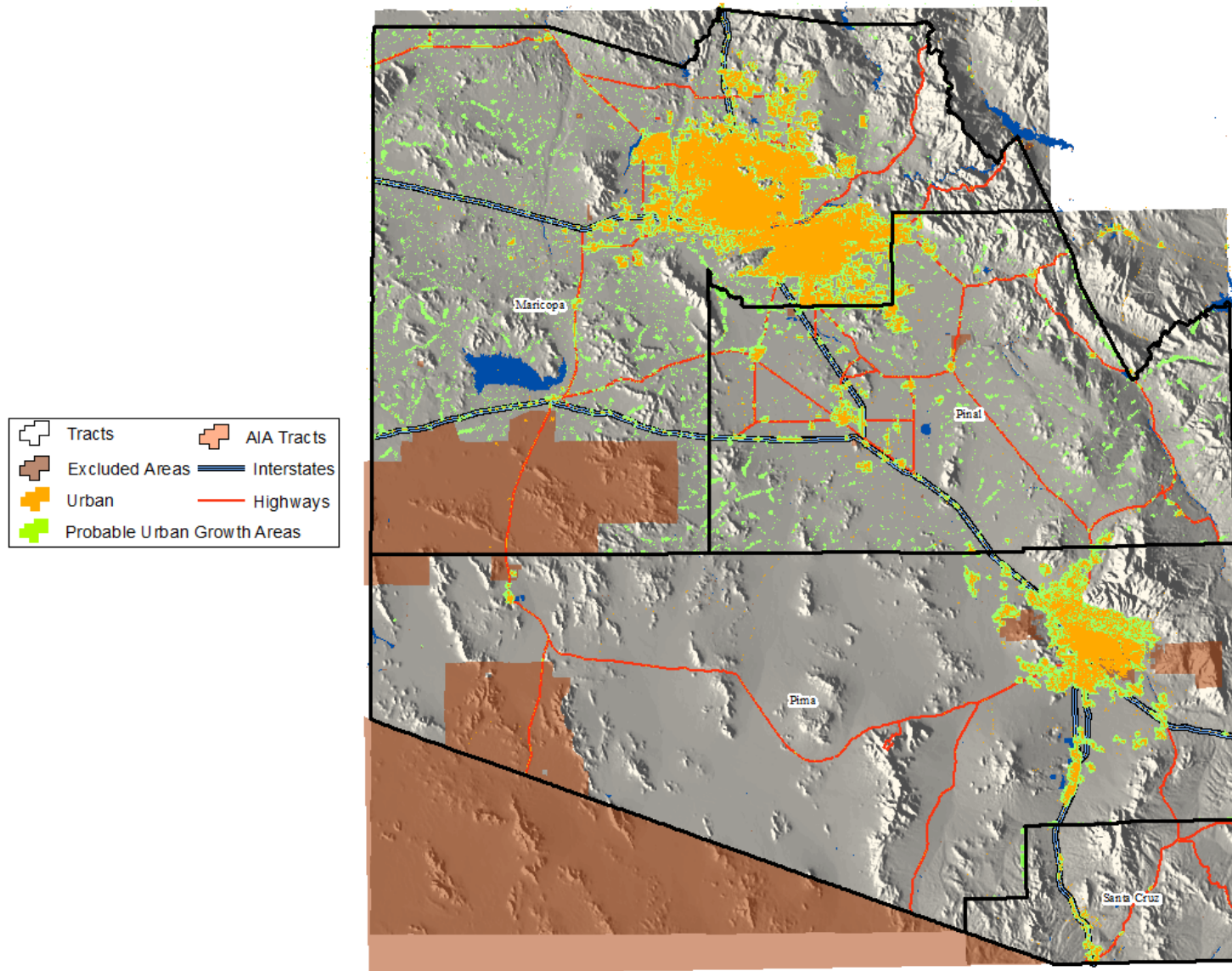


# Study Area 3: Arizona Counties



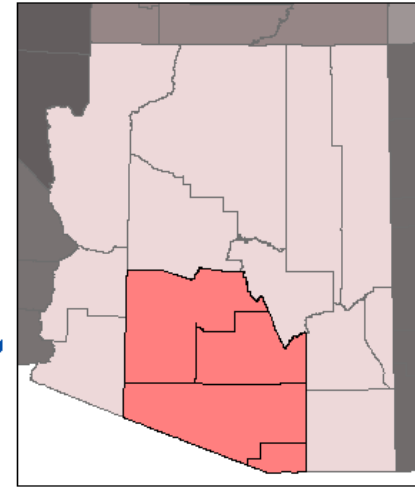
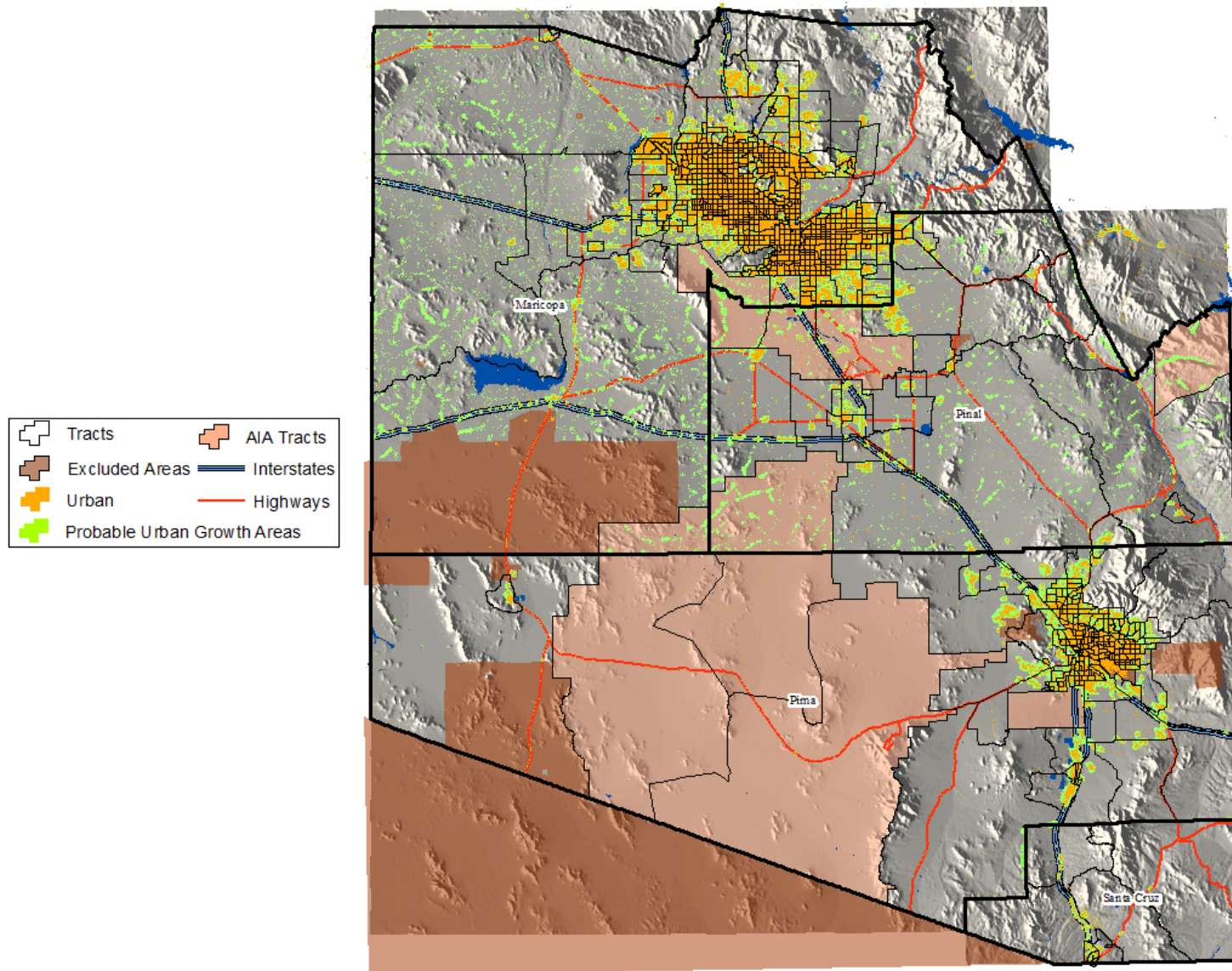


# Study Area 3: Arizona Counties

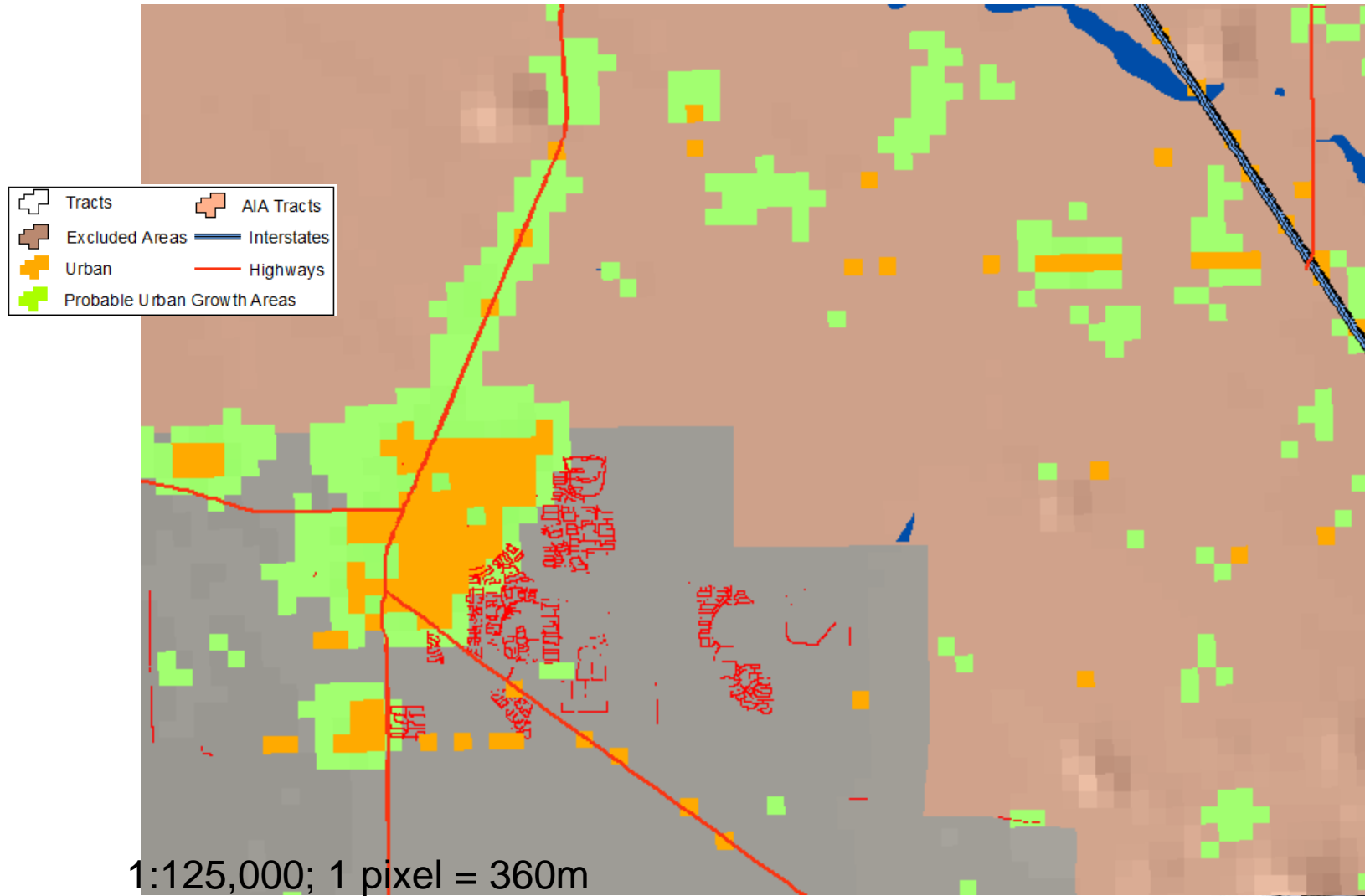




# Study Area 3: Arizona Counties



# American Indian Areas (AIA) Tracts: Model Over-predicted Growth



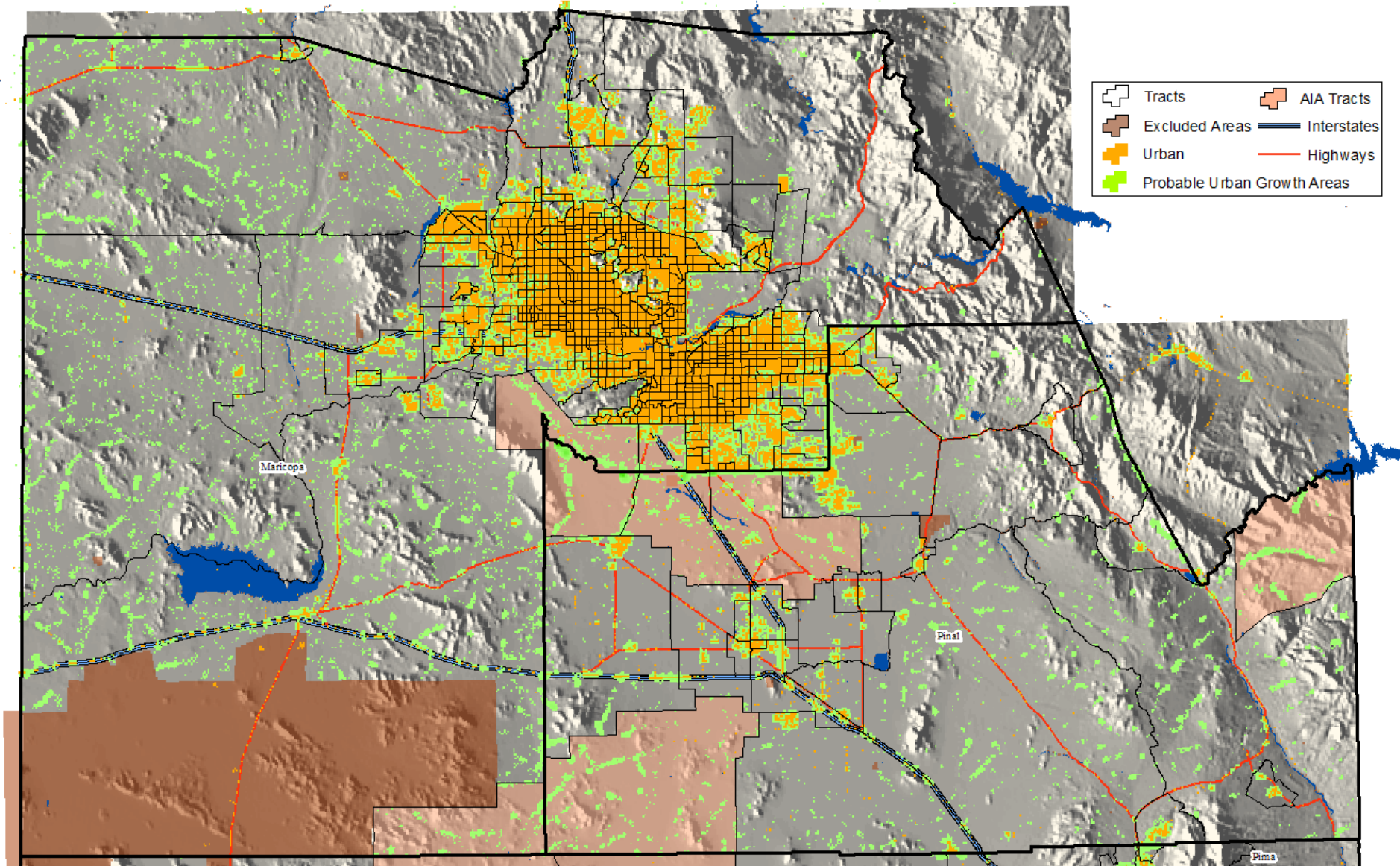
1:125,000; 1 pixel = 360m

1:125,000; 1 pixel = 360m



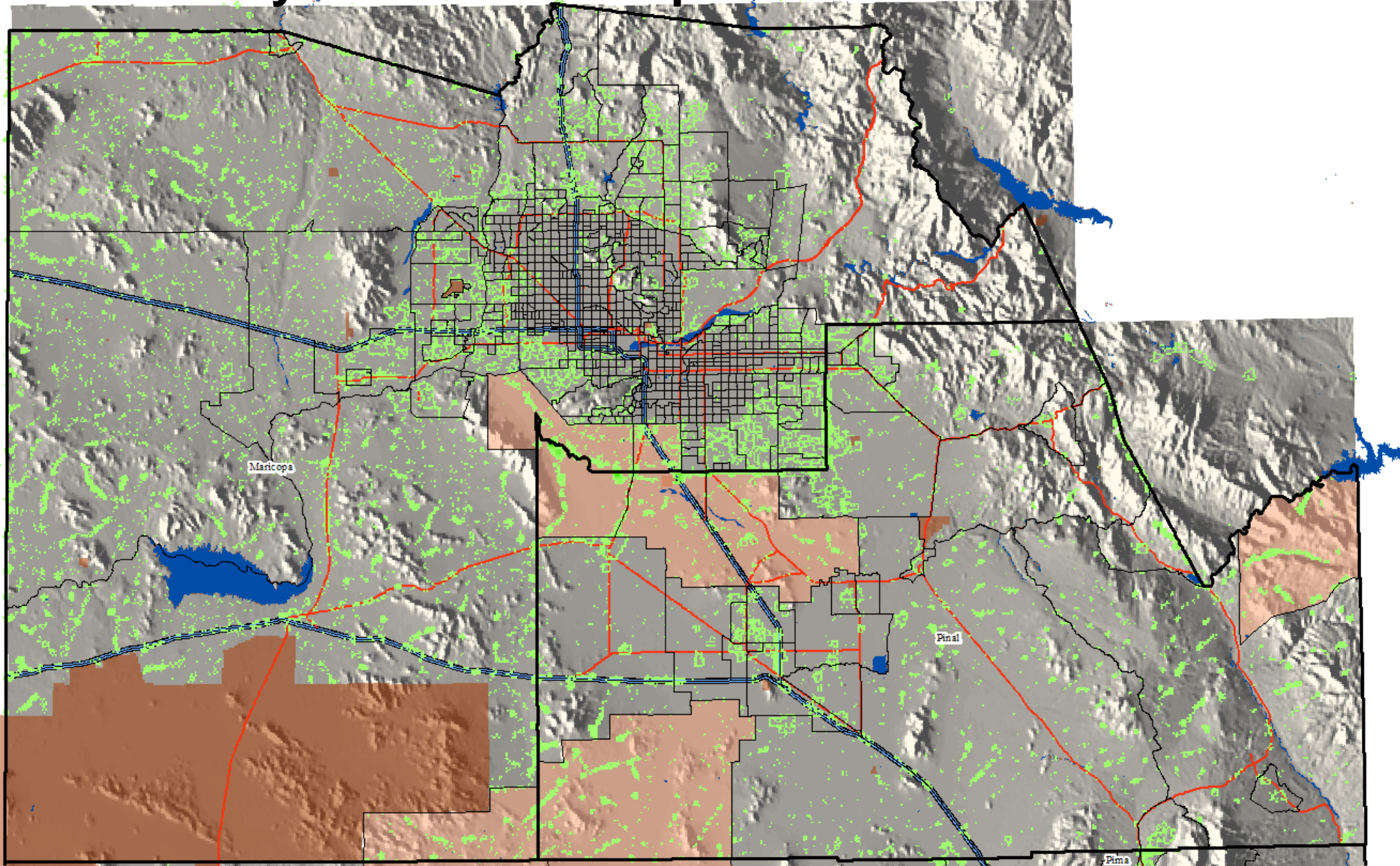


# Study Area 3: Maricopa and Pinal Counties



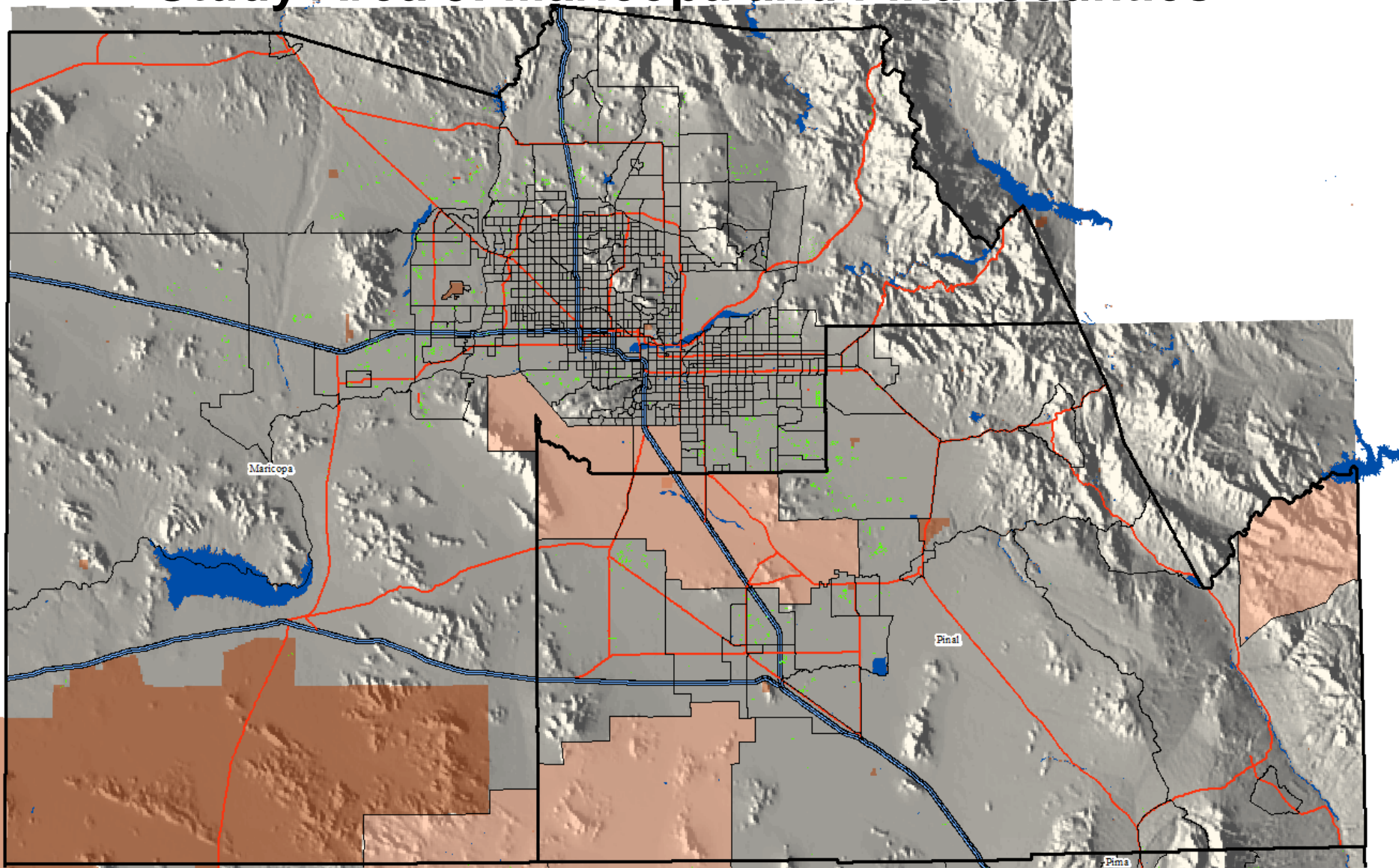


# Study Area 3: Maricopa and Pinal Counties

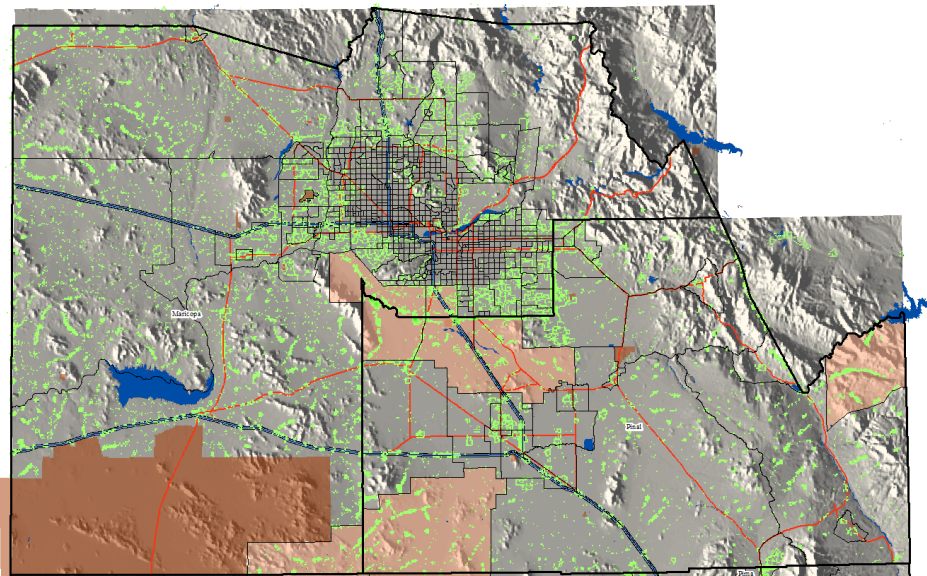




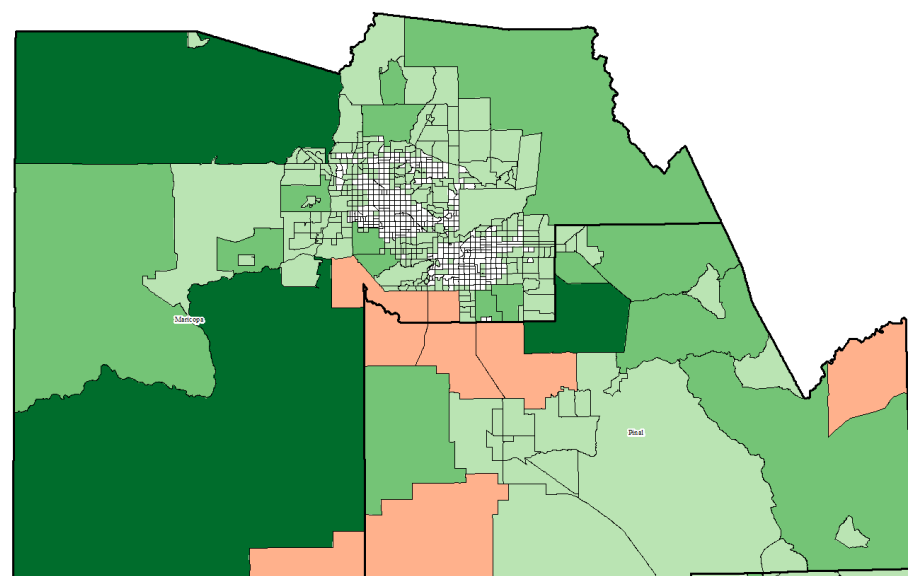
# Study Area 3: Maricopa and Pinal Counties



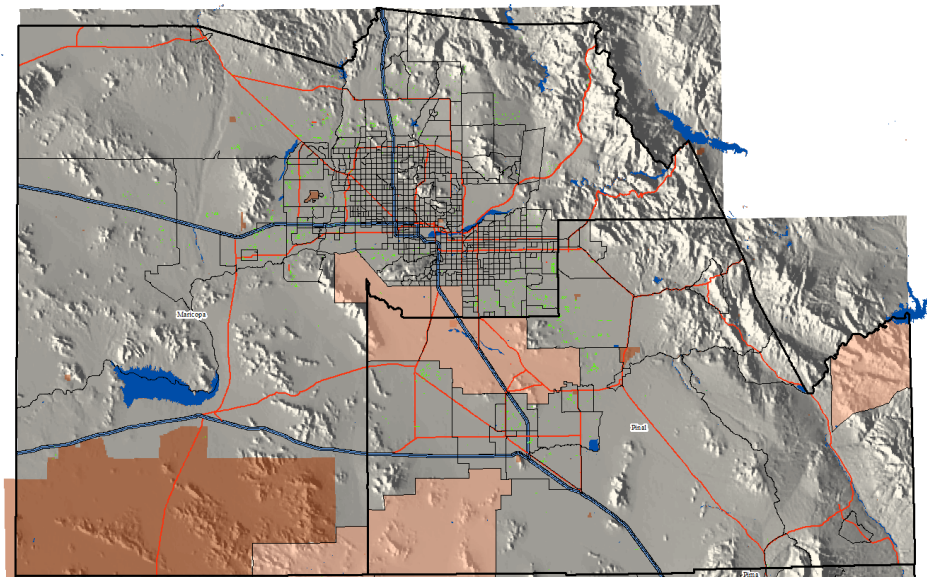




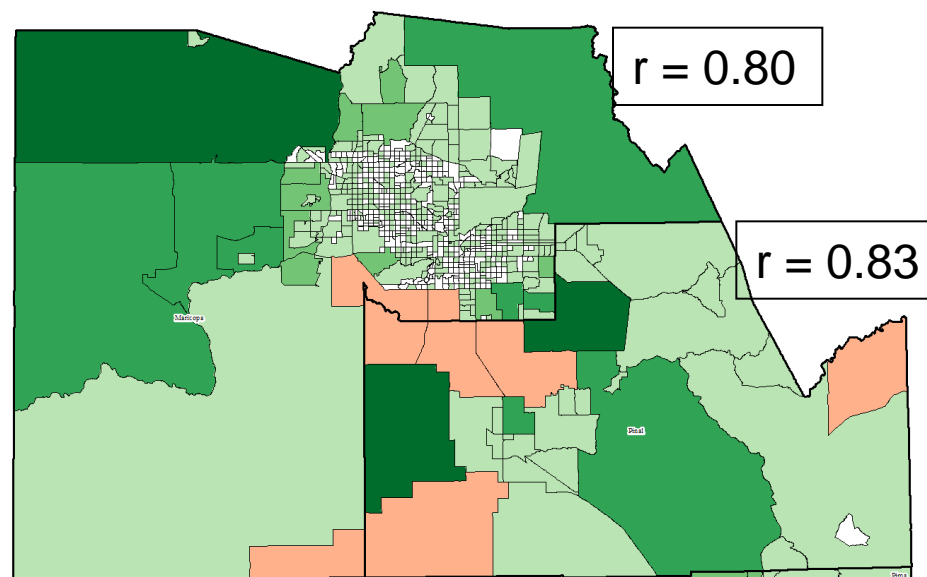
**SLEUTH Probable Growth Areas**



**SLEUTH Summed by Tract**



**Actual Road Growth**

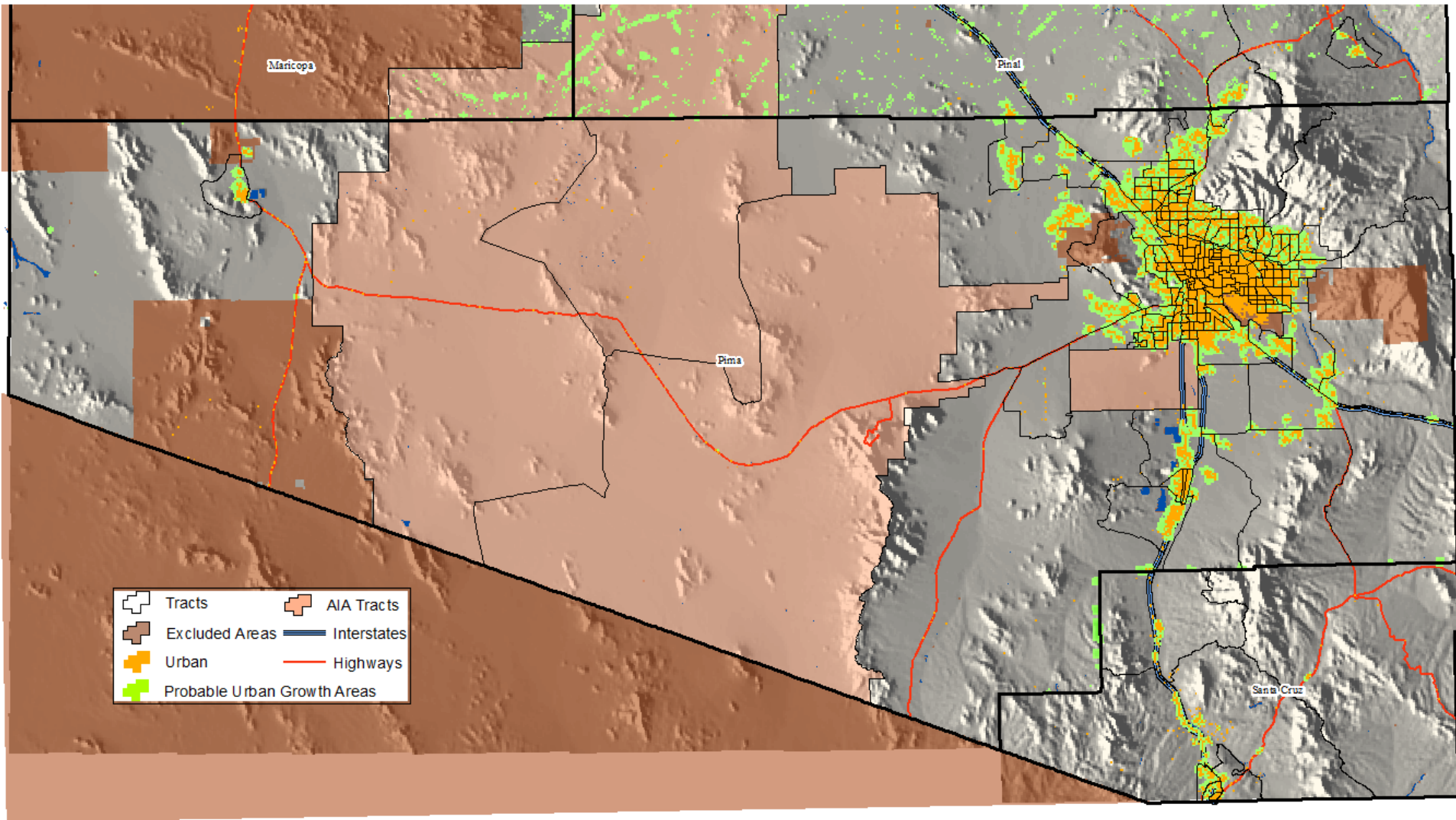


**Road Growth Summed by Tract**

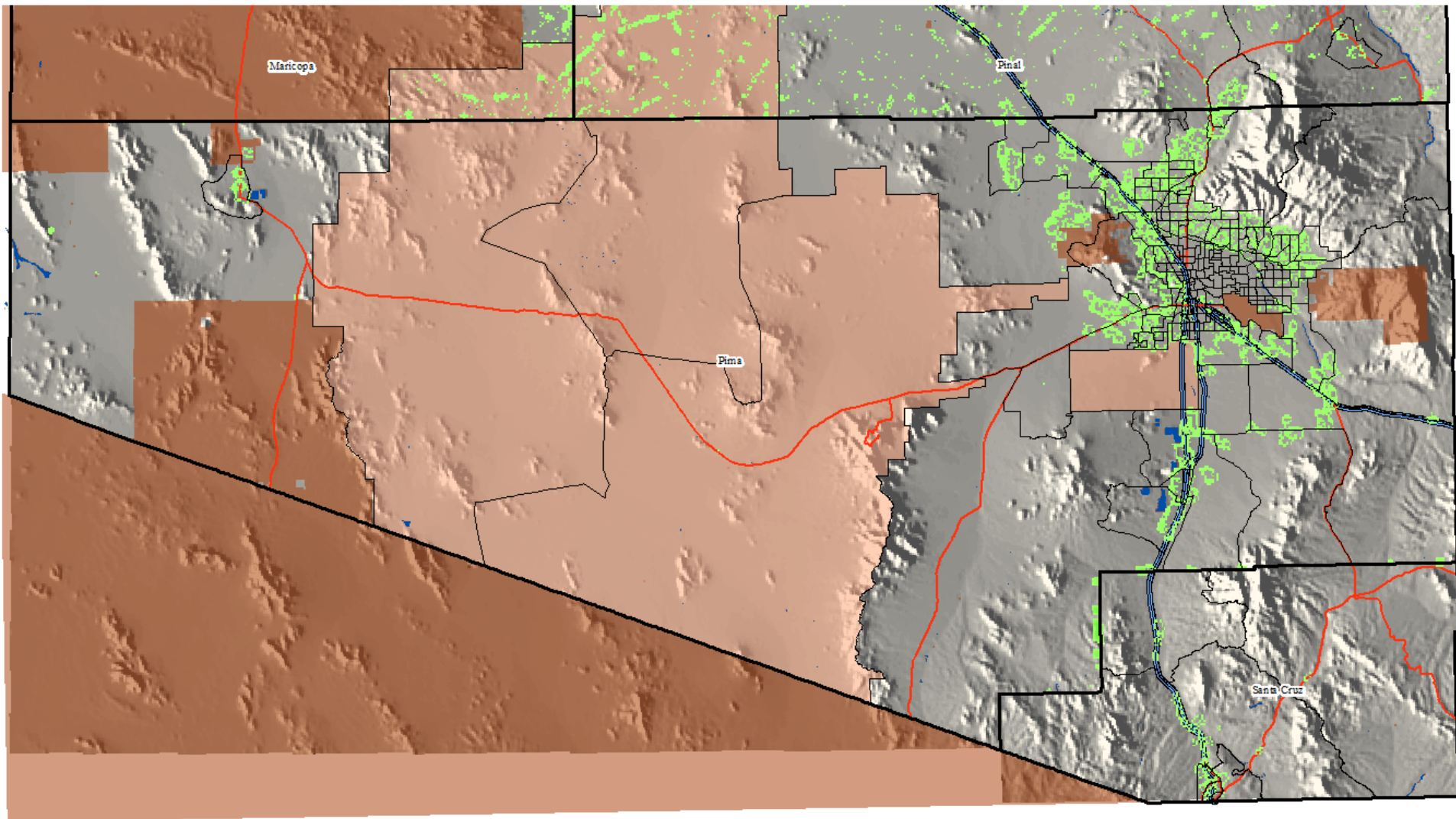




# Study Area 3: Pima and Santa Cruz Counties

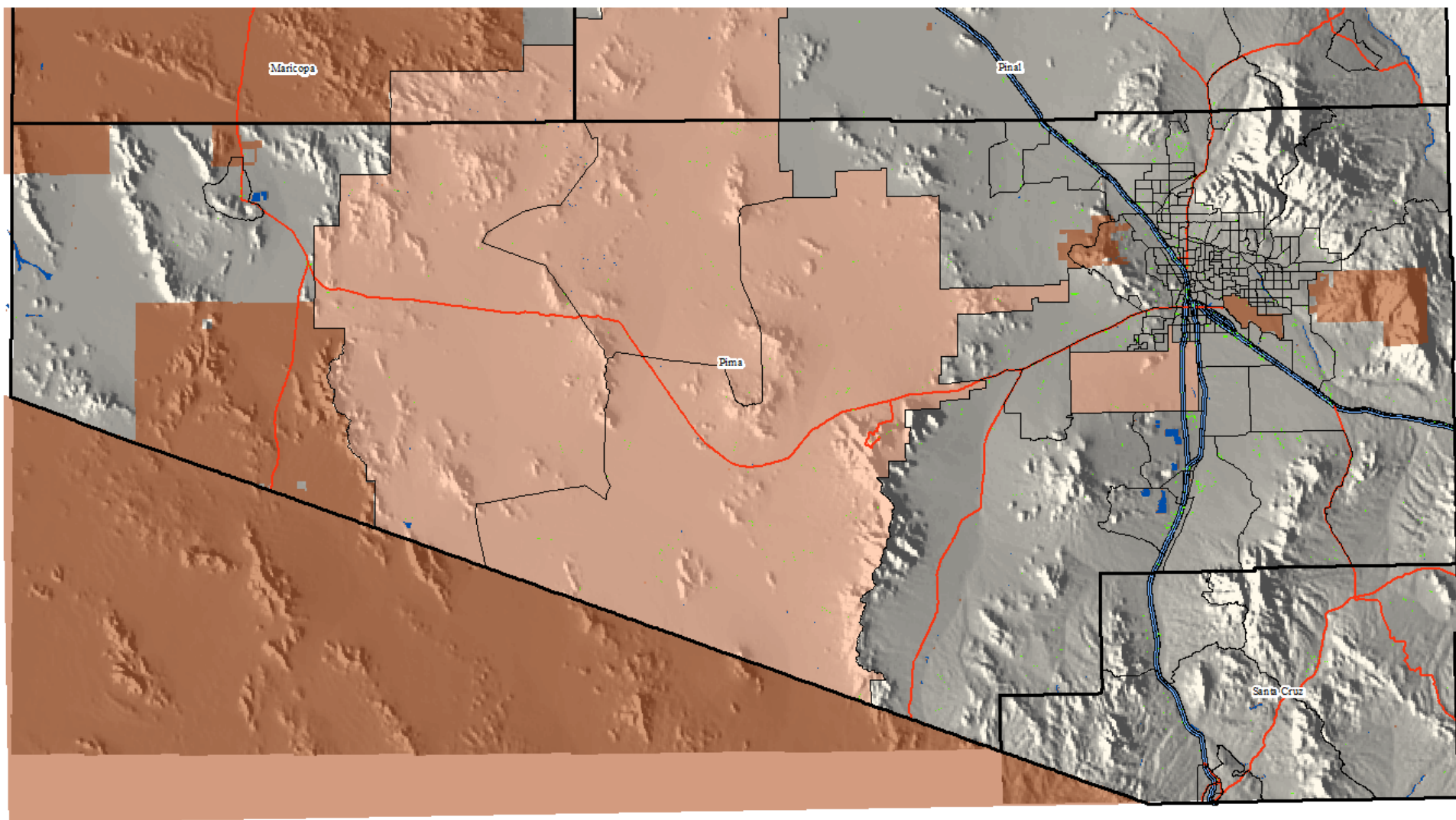


# Study Area 3: Pima and Santa Cruz Counties

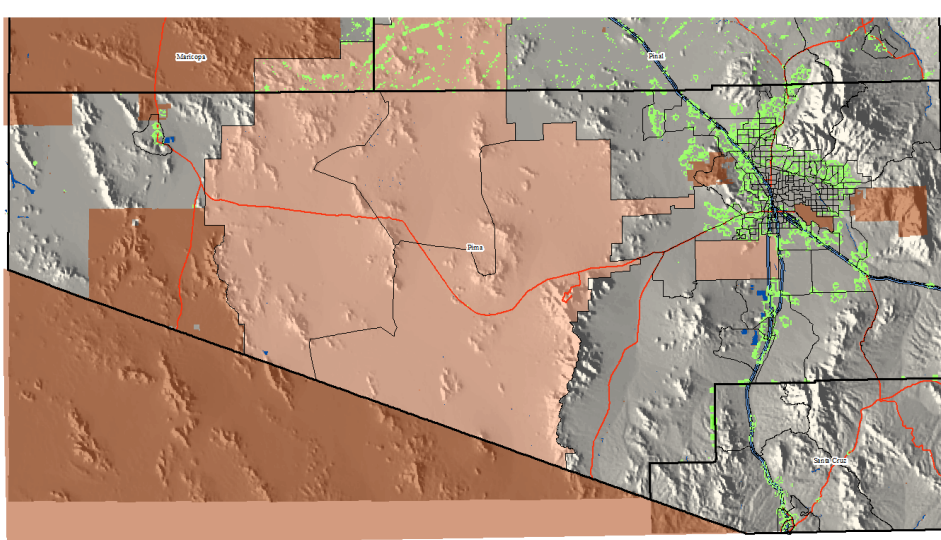




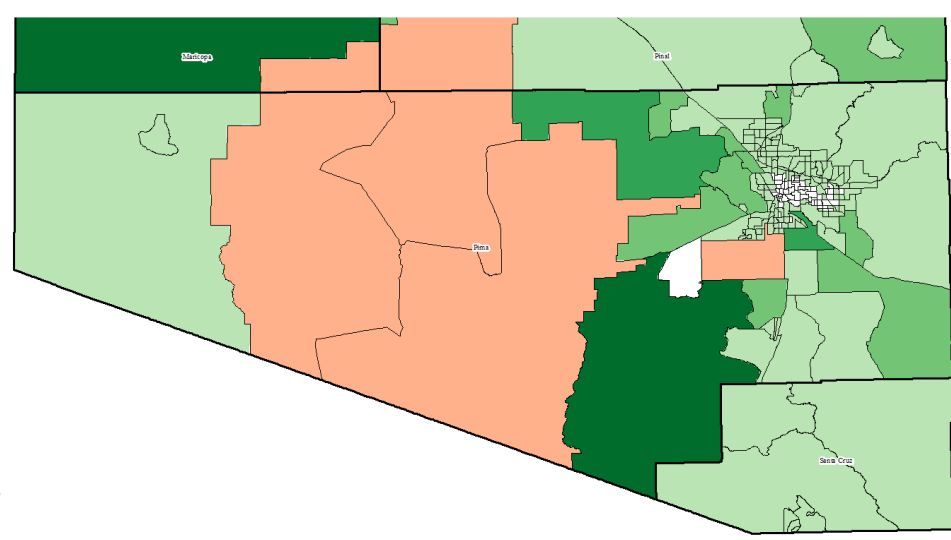
# Study Area 3: Pima and Santa Cruz Counties



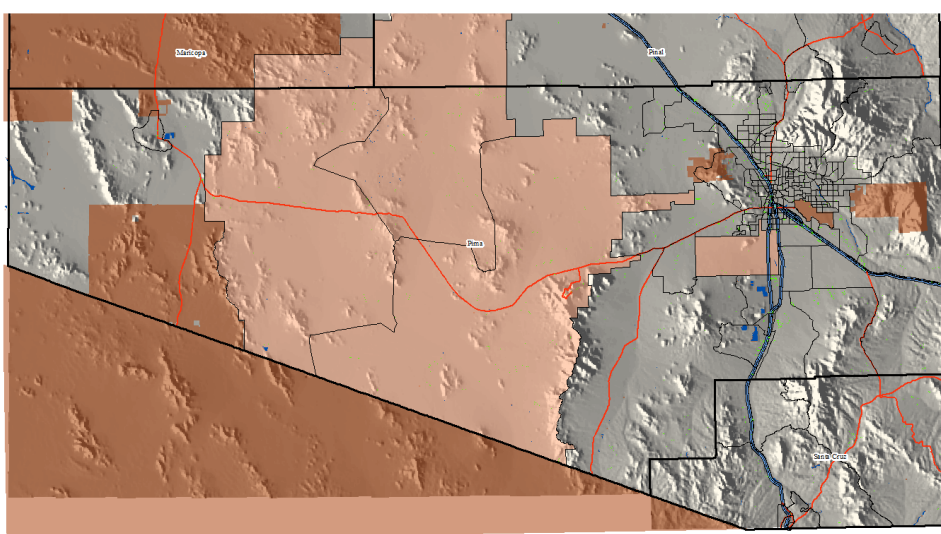




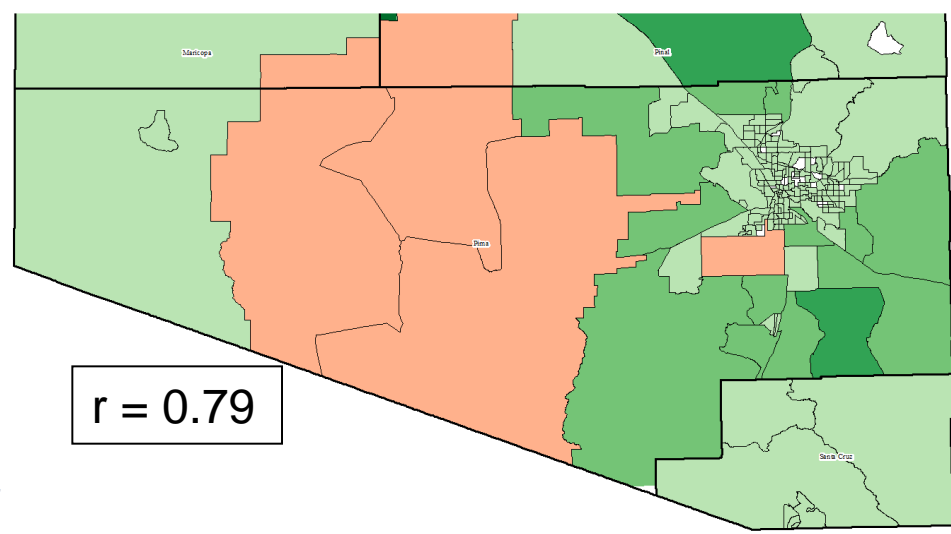
**SLEUTH Probable Growth Areas**



**SLEUTH Summed by Tract**



**Actual Road Growth**



**Road Growth Summed by Tract**

**$r = 0.15$**



	ANSI (FIPS)	Correlation Coefficients*		Area (Sq.Mi.)	# Tracts (Total)*
		Prediction to Road Growth	Area to Road Growth		
Maricopa, AZ	04013	0.80	0.53	9,224	660
Pima, AZ	04019	0.79	0.48	9,189	194
Pinal, AZ	04021	0.83	0.32	5,374	28
Santa Cruz, AZ**	04023	0.15	0.99	1,238	7
ALL ARIZONA TRACTS		0.79	0.49	25,025	889
Lincoln, SD	46083	0.99	0.11	578	4
Minnehaha, SD	46099	0.58	0.58	809	29
Rockwall, TX	48397	0.98	0.96	149	7
ALL TRACTS		0.72	0.48	26,561	929

\*Not including AIA Tracts (9400s)

\*\*Santa Cruz – Removing one outlier tract, CC becomes 0.66, but summary coefficients (ALL AZ TRACTS, ALL TRACTS) are not significantly impacted.

		Correlation Coefficient (Pearson's r)					
		Prediction to Road Growth		Area to Road Growth		# Tracts (Total)	# AIA Tracts
		excl. AIAs	incl. AIAs	excl. AIAs	incl. AIAs		
Maricopa, AZ	04013	0.80	0.79	0.53	0.53	663	3
Pima, AZ	04019	0.79	0.55	0.48	0.73	198	4
Pinal, AZ	04021	0.83	0.49	0.32	0.29	33	5
Santa Cruz, AZ*	04023	0.15	0.15	0.99	0.99	7	0
ARIZONA TRACTS		0.79	0.66	0.49	0.56	901	12
Cherokee, GA	13057	0.82	0.82	0.49		23	0
Lincoln, SD	46083	0.99	0.99	0.11		4	0
Minnehaha, SD	46099	0.58	0.58	0.58		29	0
Rockwall, TX	48397	0.98	0.98	0.96		7	0
ALL TRACTS		0.72	0.64	0.47	0.52	964	12

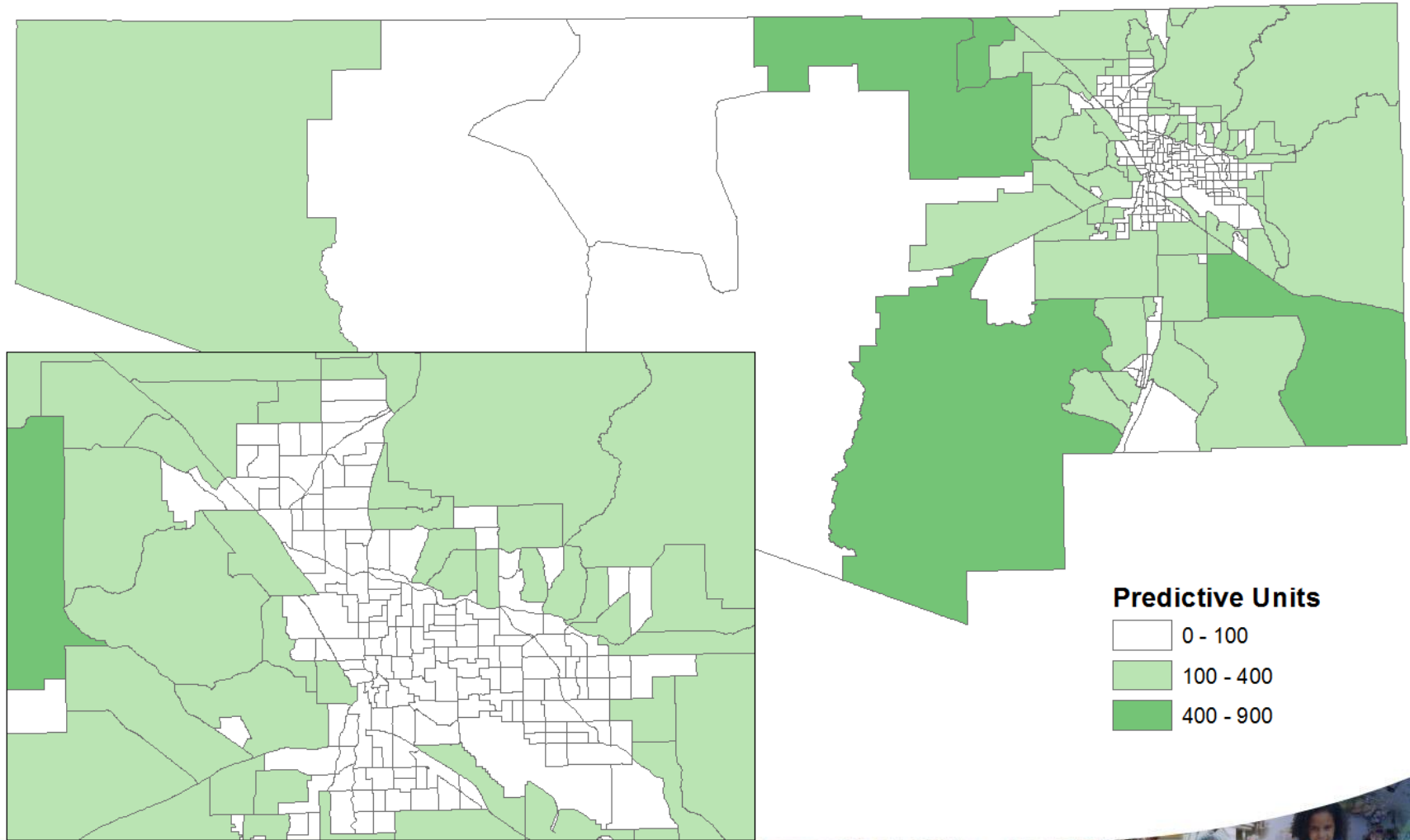
\*Santa Cruz – Removing one outlier tract, the CC becomes: 0.66

Summary coefficients (AZ TRACTS, ALL TRACTS) are not significantly impacted.



# Estimating Missing Data (Incompleteness)

## Difference between 2013 and 2011 predictions



# Lessons Learned

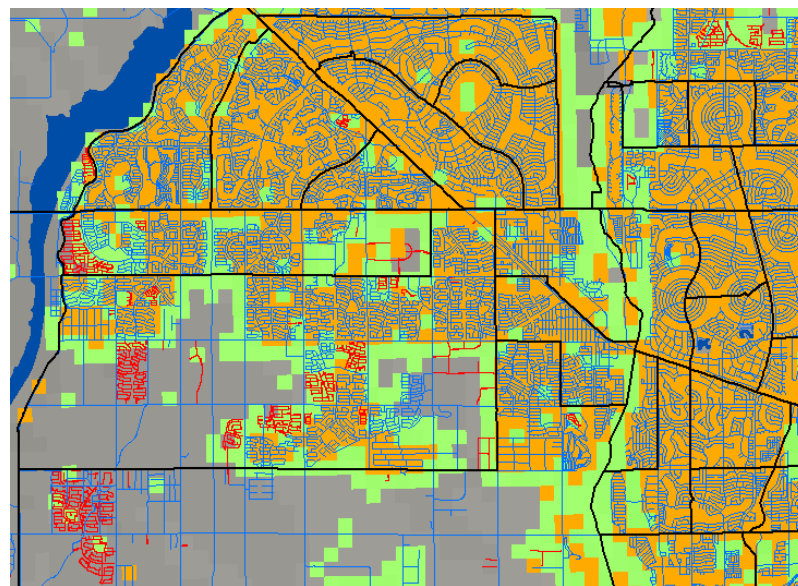
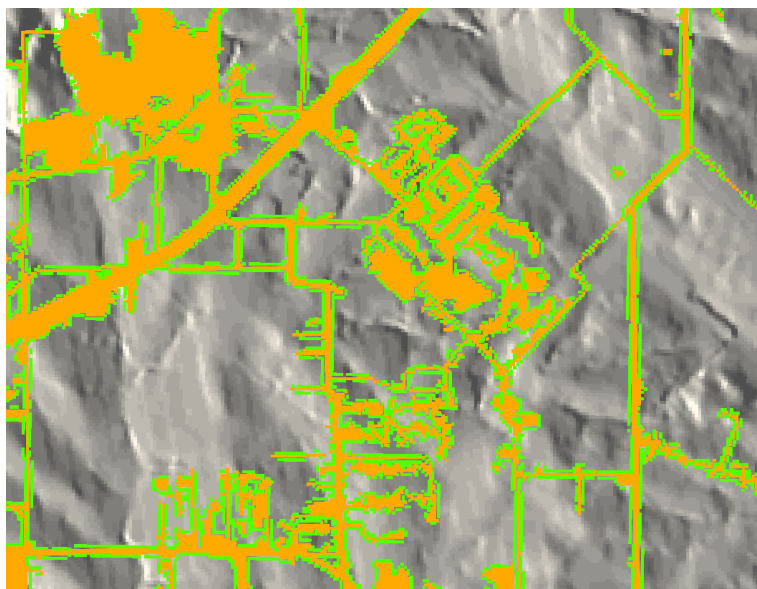
- Process Adjacent Counties Separately

## “URBAN DNA”

COUNTY	Diffusion	Breed	Spread	Slope Resistance	Road Gravity
Maricopa, AZ	75	50	1	50	1
Pima, AZ	25	50	1	25	75
Pinal, AZ	1	50	1	25	100
Santa Cruz, AZ	1	1	1	75	75
Lincoln, SD	1	1	1	1	75
Minnehaha, SD	1	1	1	1	50
Rockwall, TX	1	100	1	75	25

# Lessons Learned

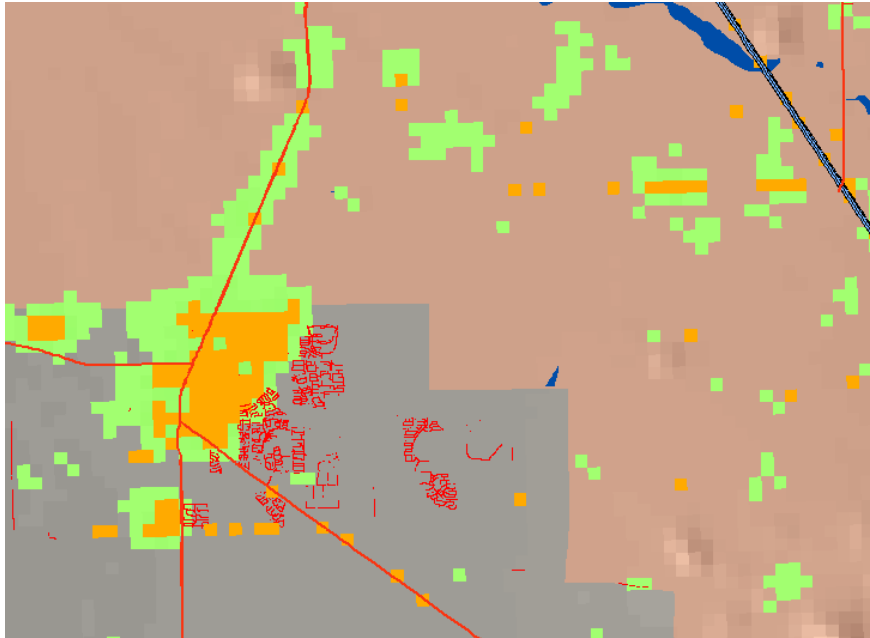
- Process Adjacent Counties Separately
- Scale: Using Larger Cells Solve Several Problems:
  - Meaning of output more applicable to the problem
  - Less processing time





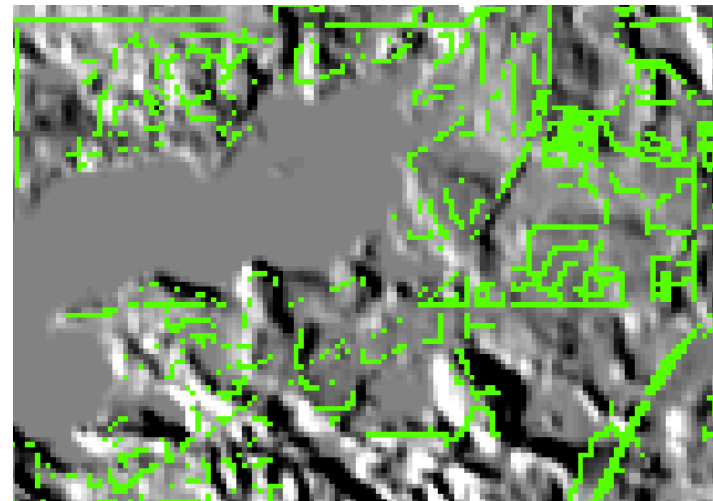
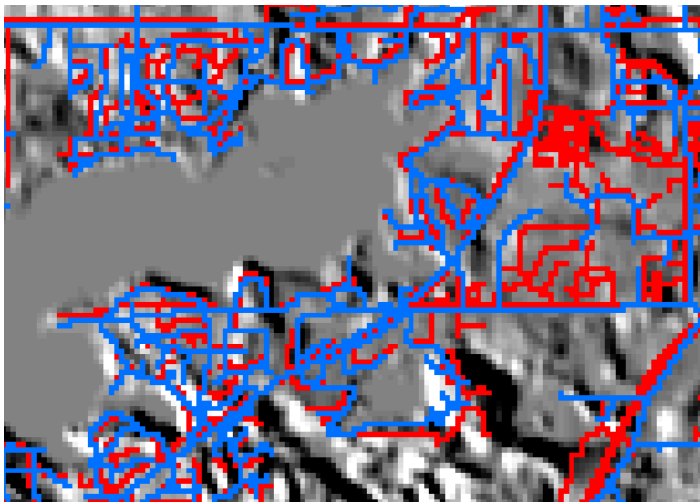
# Lessons Learned

- Process Adjacent Counties Separately
- Scale: Using Larger Cells Solve Several Problems
- Model Overpredicts Growth on AIA Tracts



# Lessons Learned

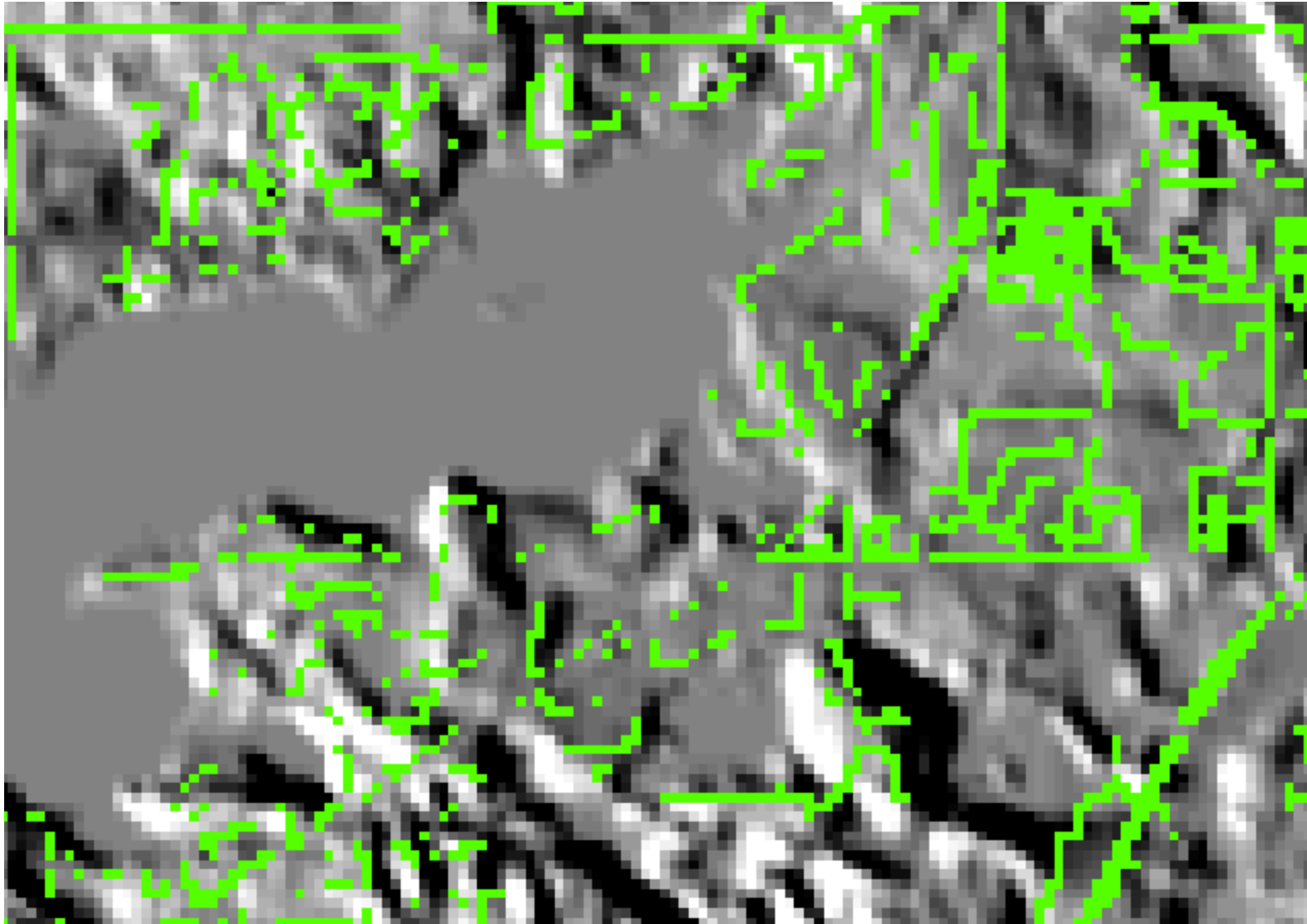
- Process Adjacent Counties Separately
- Scale: Using Larger Cells Solve Several Problems
- Model Overpredicts Growth on AIA Tracts
- Actual Growth can be Hard to Measure







Difference: ([2011] – [2006])



# Limitations

- **SLEUTH does not consider the underlying causes of urban growth, such as:**
  - Population Growth
  - Economic Growth
- **Instead focuses on analyzing and extrapolating urban growth pattern (“Urban DNA”)**
  - Diffusion
  - Breed
  - Spread
  - Slope Resistance
  - Road Gravity



# Future Possibilities

- Adapt model to constrain the outputs to match population or economic growth projections
- Adapt the model to make use of demographic inputs
- New NLCD data (2011) scheduled for release in December 2013
- Updated projections for the rest of this decade
- Imagery for specific areas could be processed to create more frequent land cover datasets with which to update predictions.





# Future Possibilities

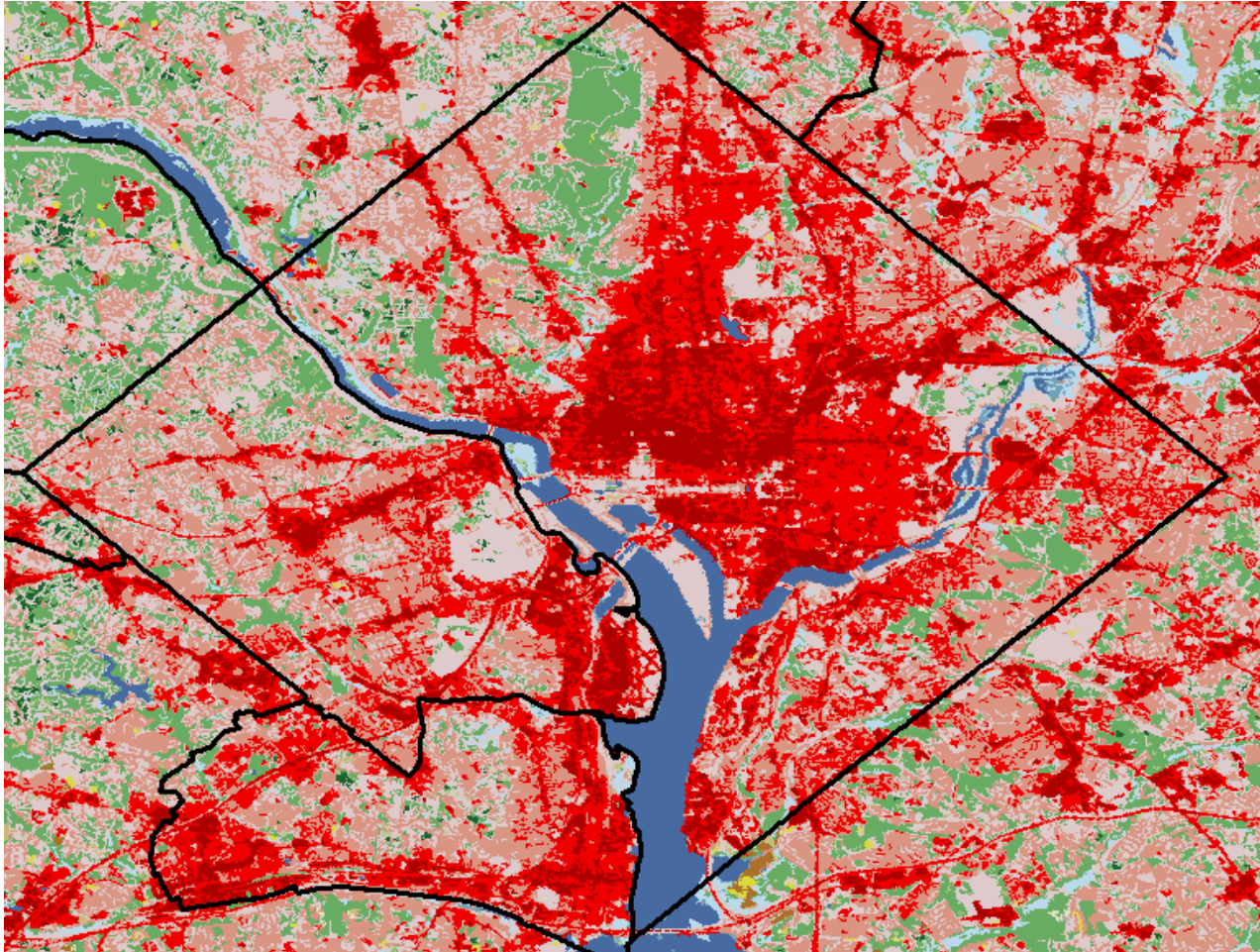
It could be useful to model urban growth for the entire country:

- SLEUTH's creator, Keith C. Clarke (UCSB), has said that he would like to see the model used for the entire United States (Clarke, 2008 and 2011).
- The 2009 study by Jantz et al. (Shippensburg University) of the entire Chesapeake Bay watershed (208 counties) remains the largest application of SLEUTH to date.
- An eventual nationwide simulation could provide estimates of completeness of coverage for TIGER that could support the Census Bureau's stated goals for targeted update operations.

# Questions? Comments?

tiernan.erickson@gmail.com

tiernan.erickson@census.gov



# References

- Clarke, K. C. (2008). A Decade of Cellular Urban Modeling with SLEUTH: Unresolved Issues and Problems. In Brail, R. K. (Ed.) Planning Support Systems for Cities and Regions (pp. 47-60). Cambridge, MA: Lincoln Institute of Land Policy.
- Clarke, K. C. (2011). "Pre-workshop comments" UGEC-NASA Workshop: Forecasting Urban Growth (FORE). Retrieved May 9, 2012, from [http://urbangrowth.ugecproject.org/index.php?title=FORE\\_Background\\_papers](http://urbangrowth.ugecproject.org/index.php?title=FORE_Background_papers).
- Clarke, K. C., and Gaydos, L. J. (1998). Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12, 699-714.
- Clarke, K. C., Hoppen, S., and Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, 24, 247-261.
- Cutsinger, J. and Galster, G. (2006). There is no Sprawl Syndrome: A New Typology of Metropolitan Land Use Patterns. *Urban Geography*, 27:3, 228-252.
- Jantz, C. A. (2009). Simulating Urban Growth with the SLEUTH Model: A Training Manual. Center for Land Use, Shippensburg University.
- Jantz, C. A., and Goetz, S. J. (2005). Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science*, 19:2, 217-241.
- Jantz, C. A. et al. (2009). Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment and Urban Systems*.
- O'Sullivan, et. al. (2012). In Heppenstall, A. J., et al. (Eds.) *Agent-Based Models of Geographical Systems* (pp. 109-123). Dordrecht, Netherlands: Springer.
- Pinto, N. N., and Antunes, A. P. (2007). Modeling and urban studies: an introduction. *ACE: Architecture, City and Environment*, 2:4, 471-485.
- Poelmans, L., and Van Rompaey, A. (2010). Complexity and performance of urban expansion models. *Computers, Environment and Urban Systems*, 34, 17-27.
- Santé, I., et al. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning* 96, 108-122.
- Silva, E.A., and Clarke, K. C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, 26, 525-552.
- Zhao, F., and Chung, S. (2006). "A Study of Alternative Land Use Forecasting Models" Report No. BD015-10, prepared for Florida Dept. of Transportation (FDOT) Systems Planning Office. Retrieved May 9, 2012 from <http://trid.trb.org/view.aspx?id=784301>.

